DOCUMENT RESUME

ED 363 625	TM 020 623

TITLE	Advances in Education Research. Volume 1, No. 1., Summer 1993.
INSTITUTION	Office of Educational Research and Improvement (ED), Washington, DC. Office of Research.
REPORT NO	OR-93-3108
PUB DATE	93
NOTE	296p.
PUB TYPE	Collected Works - General (020) -- Collected Works - Serials (022)
JOURNAL CIT	Advances in Education Research; v1 n1 Sum 1993
EDRS PRICE	MF01/PC12 Plus Postage.
DESCRIPTORS	Academic Achievement; *Adolescents; *Children; *Community Involvement; Compensatory Education; *Disadvantaged Youth; Educational Change; *Educationally Disadvantaged; Educational Policy; Educational Practices; *Educational Research; Elementary Secondary Education; Program Improvement; Remedial Instruction; Scholarly Journals; Track System (Education); Urban Schools
IDENTIFIERS	*Policy Issues

ABSTRACT

Thirteen previously published articles from selected journals are presented, which concentrate on educationally disadvantaged children and youth, and are grouped under the four general themes or categories of children and youth, school practices, community involvement, and policy issues related to the Chapter 1 program. All the articles have been cited either in "Current Index to Journals in Education (CIJE)" or in "Resources in Education (RIE)." Their locater numbers starting with EJ or ED are included in this abstract. The following articles are included: (1) "Demographic Disparities of Inner-City Eighth Graders" (Samuel S. Peng, Margaret C. Wang, and Herbert J. Walberg) (EJ440492); (2) "Educational Levels of Adolescent Childbearers at First and Second Births" (Diane Scott-Jones) (EJ436971); (3) "Explaining Within-Semester Changes in Student Effort in Junior High School and Senior High School Courses" (Douglas J. Mac Iver, Deborah J. Stipek, and Denise H. Daniels) (EJ436880); (4) "Preventing Early Reading Failure with One-to-One Tutoring: A Review of Five Programs" (Barbara A. Wasik and Robert E. Slavin) (ED324122); (5) "Responsive Practices in the Middle Grades: Teacher Teams, Advisory Groups, Remedial Instruction, and School Transition Programs" (Douglas J. Mac Iver and Joyce L. Epstein) (EJ436976); (6) "School Competency Testing Reforms and Student Achievement: Exploring a National Perspective" (Linda F. Winfield) (EJ415877); (7) "Achievement Effects of Ability Grouping in Secondary Schools: A Best-Evidence Synthesis" (Robert E. Slavin) (EJ417571 and ED322565); (8) "The Variable Effects of High School Tracking" (Adam Gamoran) (EJ456685); (9) "Community Involvement and Disadvantaged Students: A Review" (Saundra Murray Nettles) (EJ436841); (10) "Using Community Adults as Advocates or Mentors for At-Risk Middle School Students: A Two-Year Evaluation of Project RAISE" (James A. McPartland and Saundra Murray Nettles) (EJ436975 and ED337536); (11) "Lessons from the Field: Case Studies of Evolving Schoolwide Projects" (Linda F. Winfield) (EJ438594); (12) "Modifying Chapter 1 Program Improvement Guidelines To Reward Appropriate Practices" (Robert E. Slavin and Nancy A. Madden) (EJ438596); and (13) "Chapter 1 Program Improvement: Cause for Cautious Optimism and a Call for Much More Research" (Sam Stringfield, Shelley H. Billig, and Alan Davis) (EJ438600). References accompany each article. (SLD)

OFFICE of RESEARCH

Volume 1, No. 1 Summer 1993

ED 363 625

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

BEST COPY AVAILABLE

TM020623

U.S. Department of Education
Office of Educational Research and Improvement

Advances in Education Research

Office of Research
Office of Educational Research and Improvement

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
Sharon P. Robinson
Assistant Secretary

Office of Research
Joseph C. Conaty
Acting Director

Summer 1993

The Office of Research has obtained permission from the copyright holders to reproduce certain quoted material in this report. Further reproduction of this material is prohibited without specific permission of the copyright holders. All other material contained in this report is in the public domain and may be used and reprinted without special permission; citation as to source, however, is expected.

Foreword

Advances in Education Research makes available to the public peer-reviewed, scholarly research supported in whole or in part by the Office of Educational Research and Improvement through its educational research and development programs. The goals of *Advances in Education Research* are to: bring together from diverse scholarly sources first-rate, exemplary research that relates to an important educational theme or topic; disseminate the results of funded research more widely to researchers, educators, and policymakers; serve as a forum for discussing, debating, and exchanging research results and perspectives of researchers and education practitioners; and increase public awareness of, access to, and use of high quality education research that is central and indispensable to improving and strengthening American education.

Advances in Education Research is produced by the Office of Research of the Office of Educational Research and Improvement. The mission of the Office of Research is to support research that helps to ensure equal access to education and to promote educational excellence throughout the nation. The aim is to provide to the American public the best available research-based information about every level of education. These goals are accomplished through basic and applied research carried out by the Office of Research, and by universities, school districts, teachers, and individuals across the nation. Accordingly, three questions guide the work the Office of Research supports and does: Is an important problem being addressed? Will greater knowledge or understanding come from the results of this work? What utility or benefit will this work have for education? For further information on *Advances in Education Research*, and on the Office of Research and the work it supports, or to contribute articles, please write or call:

Office of Research
555 New Jersey Avenue NW
Washington, DC 20208-5573
Voice: 202-219-2079
Fax: 202-219-2030

This first volume of *Advances in Education Research* includes previously published articles from selected refereed journals which are briefly described at the end of this volume. The articles are reproduced with the permission of the authors and the journals in which they originally appeared. They were written by one or more individuals affiliated with the Office of Research's university-based National Educational Research and Development Centers. However, the views expressed in these articles are those of the authors and do not necessarily reflect the position or policy of the Office of Educational Research and Improvement or the Office of Research, and no official endorsement should be inferred.

Contents

This volume presents selected research articles related to issues on educationally disadvantaged children and youth. Each article in this volume was previously published in a refereed journal.

Introduction I–1

Children and Youth

Demographic Disparities of Inner-City Eighth Graders
Samuel S. Peng, Margaret C. Wang, and Herbert J. Walberg II–1

Educational Levels of Adolescent Childbearers at First and Second Births
Diane Scott-Jones II–20

Explaining Within-Semester Changes in Student Effort in Junior High School and Senior High School Courses
Douglas J. Mac Iver, Deborah J. Stipek, and Denise H. Daniels II–40

School Practices

Preventing Early Reading Failure With One-to-One Tutoring: A Review of Five Programs
Barbara A. Wasik and Robert E. Slavin III–1

Responsive Practices in the Middle Grades: Teacher Teams, Advisory Groups, Remedial Instruction, and School Transition Programs
Douglas J. Mac Iver and Joyce L. Epstein III–23

School Competency Testing Reforms and Student Achievement: Exploring a National Perspective
Linda F. Winfield III–58

Achievement Effects of Ability Grouping in Secondary Schools: A Best-Evidence Synthesis
Robert E. Slavin III–75

The Variable Effects of High School Tracking
Adam Gamoran III–104

Contents

Community Involvement

Community Involvement and Disadvantaged Students: A Review
Saundra Murray Nettles IV–1

Using Community Adults as Advocates or Mentors for At-Risk Middle School Students: A Two-Year Evaluation of Project RAISE
James M. McPartland and Saundra Murray Nettles IV–28

Policy Issues: Chapter 1

Lessons From the Field: Case Studies of Evolving Schoolwide Projects
Linda F. Winfield V–1

Modifying Chapter 1 Program Improvement Guidelines to Reward Appropriate Practices
Robert E. Slavin and Nancy A. Madden V–11

Chapter 1 Program Improvement: Cause for Cautious Optimism and a Call for More Research
Sam Stringfield, Shelley H. Billig and Alan Davis V–22

Journal Descriptions VI–1

Introduction

Perhaps the most urgent and compelling issue facing American education is how best to improve and strengthen the quality of education and performance of educationally disadvantaged students—students who confront multiple kinds of problems which interfere with and impede their success in school, and which are frequently beyond their control. The articles in this issue of *Advances in Education Research* address and focus on some of the most important problems and circumstances ermane to educationally disadvantaged children and youth. These articles cover distinct—though interrelated—themes, topics, and levels of education. They reflect an interdisciplinary approach to research on educationally disadvantaged students. They also represent various conceptual, methodological, and analytical approaches. And, they demonstrate the kinds of significant contributions research can and does make to preventing and overcoming problems and barriers faced by educationally disadvantaged students.

The 13 articles included in this first volume are grouped under four general themes or categories: (1) children and youth; (2) school practices; (3) community involvement; and (4) policy issues related to Chapter 1.

Children and Youth

Within this broad category Peng, Wang, and Walberg (*Demographic Disparities of Inner-City Eighth Graders*) give us a context within which the backgrounds of inner-city students can be better understood. Based on an analysis of the National Center for Education Statistics' National Education Longitudinal Study of 1988 (NELS:88), Peng, Wang, and Walberg provide a demographic and socioeconomic profile of middle-grade students, comparing those who are enrolled in inner-city schools with those who are enrolled in schools in other types of communities. Peng and his colleagues show that children attending inner-city schools are quite different from those children attending schools in suburban or rural communities. For example, the vast majority of children in inner-city schools are African American and Hispanic, they do not live with both natural parents, and they live in poverty.

The results of Peng et al.'s analysis underscore the importance for educators of knowing and understanding more than has been traditionally required. Effective approaches to teaching and learning of inner-city youth must reflect an awareness and appreciation of students' backgrounds, readiness, motivations, interests, and developmental skills. Educators will need to have a keen sense not only of the academic strengths and weaknesses of inner-city children, but also of their cultures and family backgrounds. These findings also imply that schools alone cannot correct many of the problems affecting the education of inner-city children. That is, schools must involve family members and must work with community health and social service agencies to prevent and solve many of the problems confronting inner-city youth.

Introduction

Diane Scott-Jones (*Educational Levels of Adolescent Childbearers at First and Second Births*) demonstrates that the patterns of childbearing and educational attainment among white, black, and Hispanic adolescents, 15–19 years of age, vary significantly. For instance, young Hispanic mothers have fewer years of schooling than do their black and white counterparts. And, compared to white mothers black adolescent mothers generally complete the same, or slightly more number of grades.

Scott-Jones' analysis also shows that the number of school years completed by adolescent mothers in the 15–19 age group is, on average, lower than that of the national median age group. This is not true in all cases, particularly for 15-year-olds, and for 15–16-year-old black mothers. The median number of years of schooling completed by 15-year-old mothers is basically the same as the national median. Further, the median educational level of 15- and 16-year-old black mothers is higher than that of the national cohort.

In addition, Scott-Jones' study shows that the educational levels of fathers and mothers in the 15–19 age cohort are positively correlated within each racial/ethnic group. The educational implications of Scott-Jones' research for adolescent pregnancy are significant and far reaching. It suggests the importance of establishing effective school-based policies and programs that (a) deter premature sexual activity and unplanned pregnancies, (b) promote the educational progress of adolescents who become pregnant, and (c) prevent students from dropping out of school because of early parenthood.

The last article included in the "Children and Youth" section looks at another important issue—student effort. Mac Iver, Stipek, and Daniels (*Explaining Within-Semester Changes in Student Effort in Junior High School and Senior High School Courses*) make the point that regardless of the course taken, as the semester progresses some students lose interest and reduce their effort and other students do better and try harder. Why? What factors account for within-semester changes in student effort? The authors state: "Virtually every theory of motivation suggests that changes in ability perceptions partially determine changes in effort. Researchers have also cited changes in students' valuing of the course and changes in extrinsic pressures as determinants of effort changes."

In studying junior and senior high school students, Mac Iver, Stipek, and Daniels find that changes in students' perceived abilities (in a subject) directly affect their effort and the value they place on a particular subject matter. These results are "consistent with the claim that, by reducing the number of students who believe they are 'not good' in a subject, teachers can increase the number of students who work near their potential."

The work of Mac Iver and his colleagues suggests that a number of strategic changes may be required to improve and strengthen students' motivation and performance. For example, if principals and teachers want to raise students' confidence in their abilities in order to boost their classroom effort, then they may be required to make specific changes in curriculum and instruction, task structures, ability-grouping policies, and student evaluation practices.

School Practices

What evidence exists on the success or effectiveness of one-to-one tutorial programs? Are some programs more effective than others? Wasik and Slavin (*Preventing Early Reading Failure With One-to-One Tutoring: A Review of Five Programs*), using a "best-evidence synthesis" of 16 evaluation studies, review and compare the effectiveness of five one-to-one tutorial reading programs that have been used to improve the reading skills of first-graders who are at risk for reading failure.

Their review focuses on: (1) Reading Recovery, a preventive tutoring program developed in New Zealand and widely used in the U.S.; (2) Success for All, a comprehensive schoolwide program with a major one-to-one tutoring component for primary grade students; (3) Prevention of Learning Disabilities, a program based on a physiological view of learning and learning disorders; (4) Wallach Tutoring Program, a program targeted to first-graders in which paraprofessional tutors are used; and (5) Programmed Tutorial Reading, a highly structured first grade reading program using paraprofessionals, volunteers, or parents.

Despite the many differences among these programs—including the extent of their effectiveness—overall Wasik and Slavin's analysis shows substantial positive results of one-to-one tutoring compared to the results of traditional methods. Further, the effects of tutoring are generally lasting. Tutorial reading programs are most effective when they include many—instead of few—components of the reading process, when they emphasize the content of the reading program in addition to the delivery style (i.e., one-to-one tutoring), and when they use certified teachers rather than paraprofessionals. The authors also suggest that tutoring programs, although costly, appear to be more effective than other types of expensive intervention strategies (e.g., reduced class size) currently in use.

Besides offering many interesting findings and results, Wasik and Slavin's work raises questions and issues that policymakers should consider with respect to designing, implementing, and maintaining tutorial programs for educationally disadvantaged children. For example, how should educators decide on which tutoring program to use for children who are at risk for school failure? What type of cost/benefit formula should educators apply? What must educators do to ensure that the tutorial programs selected will have "sustaining effects"? And how much are we prepared to spend to achieve these results?

Junior high schools, or more accurately the middle-grade schools, are both major socializing institutions and critical academic "turning points" in the lives of America's young adolescents. Mac Iver and Epstein (*Responsive Practices in the Middle Grades: Teacher Teams, Advisory Groups, Remedial Instruction, and School Transition Programs*) examine the use and principals' perceived effects of (a) interdisciplinary teacher teams, (b) homeroom or group advisory periods, (c) remedial instruction programs, and (d) school transition programs—school practices that many educators believe respond to the needs of young adolescents. The authors base their analysis on data collected from a national sample of principals in public schools with grade seven. Are there substantial benefits to a school and its students if schools have group advisory periods, establish interdisciplinary teams, offer remedial instruction, and conduct programs for students' smooth transition to and from the middle grades?

Generally, the results of Mac Iver and Epstein's analysis suggest that most of the school practices they studied produce modest benefits. Mac Iver and Epstein also find that principals report a stronger school program overall when they invest heavily in interdisciplinary teacher teams to create supportive conditions for students and teachers. In addition, principals expect fewer students to leave high school before graduating when the school uses supportive advisory group activities or responsive remediation programs. Another finding from Mac Iver and Epstein's study is that extensive school transition programs reduce the number of students who have to repeat the grade immediately following the transition.

Based on their data, Mac Iver and Epstein predict that if schools conduct arranged group advisory activities weekly—as opposed to infrequently—then these schools may prevent 1 percent of their students from dropping out of high school before they graduate. They also predict that when schools provide an extra subject period during the school day to students who need coaching or remedial help, then these schools are likely to reduce their dropout rates by more than 1 percent.

The above results, as well as others from the Mac Iver and Epstein study, have important implications for the improvement of education in the middle grades. They suggest, for instance, that schools must make sure that responsive practices are implemented properly. They also imply that not all practices are equally beneficial, that different practices may require different implementation strategies, and that the best way to realize the full benefits of a practice may be to combine or mix it with some other practice(s).

Linda F. Winfield's article (*School Competency Testing Reforms and Student Achievement: Exploring a National Perspective*) tackles competency testing, another important school practice. Over the last two decades, a number of education reform initiatives have been designed to increase accountability and to improve student achievement outcomes. As Winfield indicates, attempts to improve student performance included "increasing graduation requirements and implementing assessment programs that define both standards of performance for students and standards of accountability for the educational system." Increasingly, states and local school districts have used minimum competency tests (MCT) as a principal means to achieve reform.

Winfield's exploratory study examines "the relationship between school-level minimum competency testing (MCT) programs and student reading proficiency as measured by the 1983–1984 National Assessment of Educational Progress (NAEP)." Winfield compares student-level proficiency outcomes for whites, blacks, and Hispanics after adjusting for selected individual and school-level differences for the 4th-, 8th-, and 11th-grade NAEP samples.

Do MCT programs make a difference? Does student proficiency improve as a result of these programs? The results of Winfield's investigation generally show a "higher level of reading proficiency for students in grades 8 and 11 attending schools with MCT programs compared with their counterparts in schools without such programs." MCT programs, however, seem to have no effect for grade 4 students.

Winfield's work brings forward important policy issues and questions, including: how and under what conditions should educators use testing to improve performance and reform schools? What care and caution must be taken when using these kinds of tests? How should NAEP, or any other type of national assessment effort, be used by educators to make or influence policy? What guidelines should there be on the use of NAEP as an instrument for school reform? Who should establish these "rules of use"? How should they be enforced? These are only a few of the policy concerns that Winfield's study raises.

Ability grouping has remained as one of the most controversial issues in education. Against this backdrop, Slavin (*Achievement Effects of Ability Grouping in Secondary Schools: A Best-Evidence Synthesis*) presents a comprehensive review of research that has evaluated the effects of ability grouping on student achievement in secondary schools. Slavin reviews 6 randomized experiments, 9 matched experiments, and 14 correlational studies that compared ability grouping to heterogeneous plans covering periods of from one semester to five years. Ability grouping is "any school or classroom organization plan that is intended to reduce the heterogeneity of instructional groups; in between-class ability grouping the heterogeneity of each class for a given subject is reduced, and in within-class grouping the heterogeneity of groups within the class (e.g., reading groups) is reduced."

Slavin's "best-evidence synthesis" review indicates that the effects of ability grouping on student achievement—as measured by standardized tests—are essentially zero at all grade levels. Slavin also concludes that (1) various models of ability grouping are equally ineffective, (2) ability grouping is equally ineffective in all subjects, notwithstanding the possible negative effect of ability grouping in social studies, and (3) there are no consistent positive or negative effects on students of high, average, or low ability who are assigned to different levels of the same course.

The results of Slavin's analysis pose a fundamental policy question many educators must confront and resolve: What justifies any form of ability grouping when the evidence shows that ability grouping has very little—if any—effect on student achievement?

Unlike Slavin, Adam Gamoran (*The Variable Effects of High School Tracking*) focuses his analysis on variation among types of tracking, not on the presence or absence of tracking or other types of ability-grouping practices. Gamoran states that the "effects of tracking in high schools depend in part on the way tracking is organized: To the extent that the structure of tracking varies across schools, tracking's impact on achievement also varies." Using data from the National Center for Education Statistics' High School and Beyond (HS&B), a national survey of high schools and their students, Gamoran examines four structural characteristics of tracking systems: (1) "selectivity"—the degree of homogeneity within tracks; (2) "electivity"—whether students choose or are assigned to track positions; (3) "inclusiveness"—the subsequent educational opportunities available to students; and (4) "scope"—the breadth and flexibility of track assignments.

Introduction

The results of Gamoran's analysis point to the significant differences among schools in the magnitude of track effects on mathematics achievement, and in net average achievement on both verbal and mathematics tests. Schools with more mobility in their tracking systems (i.e., allowing student movement from one track position to another) produce higher math achievement overall. They also have smaller gaps between tracks in both math and verbal achievement when compared to schools with more rigid tracking systems. Moderately inclusive systems (i.e., with relatively more students assigned to the college preparatory program) also have less between-track inequality in math. Furthermore, overall school achievement tends to rise in both subjects as inclusiveness increases.

With respect to differences between Catholic and public schools, Gamoran finds that Catholic schools have less inequality between tracks and higher productivity overall than do public schools, especially for math achievement. Gamoran attributes these Catholic school advantages partly to the way Catholic schools implement tracking.

These findings, as well as others, are outcomes of Gamoran's rich analysis. They show that the issue of tracking—and its effects—is detailed, varied, and complex. Gamoran's work raises a number of basic education policy questions. For instance, should there be any kind of tracking in schools? Or, should all forms of school tracking be eliminated? On what bases should decisions be made to track or not to track students? If tracking systems are to exist, what types of systems should schools implement? How should they be structured? How should students be assigned? How much homogeneity, flexibility, mobility, et cetera, should there be within and between tracks? Certainly, educators must consider and weigh the costs and benefits of tracking/not tracking, particularly in terms of maximizing overall academic achievement, and at the same time, maximizing access to learning opportunities for all students.

Community Involvement

Community participation is seen as a critical component in efforts to help prevent and solve many of the problems of educationally disadvantaged students. Saundra Murray Nettles (*Community Involvement and Disadvantaged Students: A Review*) addresses the effects of community involvement on students who encounter multiple barriers and difficulties to success in schools. The first part of Nettles' article (1) defines community involvement as "the actions that organizations and individuals (e.g., parents, businesses, universities, social service agencies, and the media) take to promote student development"; and (2) describes and conceptualizes involvement in terms of four change processes—conversion, mobilization, allocation of resources, and instruction.

Conversion is the process of bringing students from one belief, or behavioral stance, to another. Mobilization involves more active participation of people and organizations in the education process. Allocation refers to actions community entities take to provide resources to children and youth. Instruction includes activities to help students develop intellectually or learn the rules and values that govern social relationships in the community.

Introduction

The second part of Nettles' article examines the effects of the four kinds of involvement noted above through a review of 13 evaluation studies of academic intervention programs with significant input from community entities. An example of the kinds of programs evaluated is the Chicago Area Project which provided services to targeted youth, undertook community improvement, and organized activities to prevent area delinquency.

In general, Nettles shows from the studies reviewed that community involvement does have positive effects on school-related behaviors and achievement, on student attitudes, and on risk-taking behavior. There are positive outcomes for school attendance, persistence in school, pregnancy prevention, and attitudes toward school. The effects range from small to substantial overall. Nettles also indicates that the pattern of outcomes varies by community involvement type. There is an overall pattern of positive effects for programs that are classified as "allocation" or "instruction"; however, there is a mixed pattern for those programs that combine the two types.

Nettles' work is important to educators and policymakers in a number of ways. At the very least, it demonstrates that the general call for greater community involvement is well founded and worthy of support. It also reveals that the type and form of involvement may be the most important factors in achieving success.

Whereas the Nettles article above reviews evaluation studies across 13 programs, the McPartland and Nettles article (*Using Community Adults as Advocates or Mentors for At-Risk Middle School Students: A Two-Year Evaluation of Project RAISE*) examines and evaluates the effects on selected student outcomes of a single project—Project RAISE, a multifaceted approach featuring outside adults as school-based advocates and one-on-one mentors for at-risk students at seven middle schools.

McPartland and Nettles find that after 2 years in operation, Project RAISE has positive effects on improving student attendance and report card grades in English, but not on promotion rates or standardized test scores. The authors point out that the effects, though sizable, were not enough to neutralize the academic risks with which students entered the program. They find that the positive results were primarily accounted for by three of the seven sites evaluated.

McPartland and Nettles indicate that "some evidence supported interpretations that, although strong one-on-one mentoring is not an essential component of an effective program that uses outside adults to assist at-risk middle school students, the RAISE model is much more likely to show positive effects when one-on-one mentoring has been strongly implemented." They also state that success may depend as well on the size and composition of the group of students served.

The research work of McPartland and Nettles raises the following questions: what must educators do to (1) locate and recruit mentors, and (2) facilitate ongoing successful relationships between mentors and at-risk students?

Introduction

Policy Issues: Chapter 1

Since it was first enacted in 1965, Chapter 1 (previously called Title I) has served as the cornerstone of the federal government's effort to help public schools meet the special educational needs of educationally disadvantaged children. Over the years, Chapter 1 has provided state and local school districts with about $80 billion. Chapter 1 is the largest categorical federal elementary and secondary education program, and is targeted to improving the academic achievement of at-risk students. With this as a backdrop, each article in this section focuses on some aspect(s) of this major federal program. Taken together, these articles cover a number of salient issues that should be considered by policymakers and educators as they continue their deliberations and discussions on making Chapter 1 more responsive to its intended beneficiaries—educationally disadvantaged school children.

Winfield's 2-year qualitative study (*Lessons From the Field: Case Studies of Evolving Schoolwide Projects*) describes changes that occurred in one of the nation's largest urban school systems following passage of the Hawkins-Stafford Amendments, which brought on what some consider to be the most sweeping changes in the history of Chapter 1 legislation. The Hawkins-Stafford Amendments, as Winfield indicates, allow schools to use Chapter 1 funds for schoolwide projects (SWPs) when 75 percent or more of the students in these schools are economically disadvantaged. A major goal of the amendments is to upgrade and improve the entire school program and to minimize administrative and instructional program fragmentation. Winfield uses a case study method to describe the central office and system role changes at the elementary school level at 11 sites.

Winfield points out that the school system's approach to schoolwide projects involves five main features: (a) a whole-school approach based on "effective schools" research; (b) a school-based management strategy requiring school staff and parent involvement; (c) an ongoing monitoring process to gauge individual student, class, and school performance; (d) a district-based support system at central and subdistrict offices to provide staff and parent training; and (e) a concentration of resources so that funds beyond the minimum amounts would be committed from Chapter 1 and operating budgets.

The results of Winfield's case study reveal that while schools use Chapter 1 funds in a number of ways, nearly all schools use funds to establish an additional teaching position to lower the teacher-student ratio during math and reading instruction. The author concludes that schoolwide projects (SWPs) have the potential for improving the learning outcomes of educationally disadvantaged students. But this potential can be realized "only if adequate support for change is provided at the central office or district level and if sufficient resources are devoted to human resources and professional development."

Winfield's study raises many important questions. For example, if schools are to change from offering a traditional Chapter 1 program to implementing a more integrated one that focuses on all students, are district central offices prepared to make analogous changes? Are district offices committed to and prepared to make the necessary structural and operational changes to expand the number of schoolwide project schools in their district and to provide effective coordination and delivery of direct services to SWP schools? What will it take—and at what cost—for districts to move away from traditional bureaucratic procedures to more flexible, responsive-oriented approaches?

The article by Slavin and Madden (*Modifying Chapter 1 Program Improvement Guidelines to Reward Appropriate Practices*) looks at possible effects of Chapter 1 assessment guidelines. Slavin and Madden discuss how new accountability guidelines have helped educators to focus on the outcomes of Chapter 1 programs. However, these guidelines may also result in rewarding counterproductive practices. They may possibly hinder early intervention programs like preschool, kindergarten, and first-grade programs that increase the baseline for later school performance gains.

Moreover, Slavin and Madden say that these guidelines may reward student retentions, "which significantly increase normal curve equivalent (NCE) gains. They may also focus teaching on narrow, easily measured objectives".

Their article offers a different approach to Chapter 1 accountability that rewards schools for reducing the number of students who fail to meet minimum standards on tests that are relevant and broad-based. Students who are held back or untested would be counted as not meeting minimum standards. Services designed to improve programs would be expanded substantially and would be provided to all Chapter 1 schools.

Slavin and Madden discuss the advantages and problems in using their proposed accountability approach. For instance, the authors note that "a school undergoing major demographic changes might appear to be declining in the percentage of students meeting minimum standards. This could be dealt with by allowing schools to submit demographic data (e.g., increases in the percentage of students qualifying for free lunch) to explain any declines." Or, "it may be unfair to hold schools fully responsible for students new to the school. This problem might be solved by counting only students in the school for at least two years." These are but a few of the potential problems and solutions Slavin and Madden address in their discussion on modifying Chapter 1 improvement guidelines.

The last sentence in Slavin and Madden's article poses a most important question: what must policymakers and educators do to ensure that Chapter 1 funds "are buying the most effective programs possible and that Chapter 1 policies are rewarding school practices conducive to the success of all children"?

The final article, by Stringfield, Billig, and Davis (*Chapter 1 Program Improvement: Cause for Cautious Optimism and a Call for More Research*), provides the results of a multistate survey of schools targeted for program improvement. The authors state: "The program improvement provisions in the Hawkins-Stafford Amendments to Chapter 1 rest on the optimistic premise that school-level accountability pressures directed at Chapter 1 will lead to higher academic achievement for educationally disadvantaged students." While the legislation may be unrealistic in assuming that improvement is primarily an act of will, it does correctly focus on the school as the proper level of change.

To determine the local responses to the new Chapter 1 provisions, Stringfield and his colleagues studied the results of a survey of principals of over 200 schools identified for program improvement in three states. The researchers found that more than two-thirds of the responding schools had begun to implement programmatic changes. They report that fully 84 percent of the respondents supported the legislative provisions. The authors' results suggest that when program improvement efforts are carefully implemented, they can lead to greater understanding of the role of Chapter 1 in schools and to better staff perceptions of compensatory education. Stringfield, Billig, and Davis conclude that more research is needed to study the effects of the Chapter 1 legislation, and to provide options to low-performing schools.

The articles in this volume give the reader a better understanding of important individual, social, and institutional conditions that frequently contribute to students being at-risk of school failure, and suggest ways of correcting these conditions. At the same time, they highlight many preventive measures that could be taken to avoid placing children at-risk and to relentlessly ensure success in school. Individually and collectively, the 13 articles add to our knowledge of what really matters, what does make a difference, and what policy decisions should be made so that children from educationally disadvantaged backgrounds can triumph in school. In sum, these articles not only provide us with innovative directions for future research, but they also stimulate our thinking about a range of new and fascinating education questions and ideas—frequently overlooked—that originate from basic and applied research.

Ronald J. Pedone
Editor
Office of Research

Inner-city pupils differed from others in racial and ethnic backgrounds, family incomes, parents' education and employment, and family composition.

DEMOGRAPHIC DISPARITIES OF INNER-CITY EIGHTH GRADERS

SAMUEL S. PENG
National Center for Education Statistics
U.S. Department of Education
MARGARET C. WANG
Temple University
HERBERT J. WALBERG
University of Illinois at Chicago

Education of inner-city children is often characterized by high drop-out rates and low achievement test scores. Drop-out rates are over 40% in some cities (Hahn, Danzberger, and Lefkowitz 1987), and test scores on average are below the national norm or ranked the lowest among children in different communities (Ornstein and Levine 1989; National Center for Education Statistics 1990). These problems persist despite tremendous improvement efforts devoted to schools. In fact, the problems in some cities have intensified (Council of Chief State School Officers 1988).

AUTHORS' NOTE: *The research reported herein was supported in part by the Temple University Center for Research in Human Development and Education and in part by the Office of Educational Research and Improvement, U.S. Department of Education, under grant no. 30-1992-121. However, the opinions expressed in this report are those of the authors. No official endorsement is intended, nor should it be inferred. The authors wish to thank Richard Hoenes for his assistance in computer work. This article was developed when the first author was serving as a senior research associate at the Temple University Center for Research in Human Development and Education under the Interagency Personnel Act program of the U.S. Department of Education.*

URBAN EDUCATION, Vol. 26 No. 4, January 1992 441-459
© 1992 Sage Publications, Inc.

Reprinted by permission of Corwin Press, Inc.

Understanding the nature of the problems and searching for solutions continue to be educational research priorities. These priorities are reflected in the recent establishment of the National Center for Education in the Inner Cities at Temple University, funded by the Office of Educational Research and Improvement of the U.S. Department of Education.

One major source of the education problems in the inner city is rooted in the students' demographic and socioeconomic backgrounds. Previous studies have shown that what students bring to schools greatly determines the difference among schools (e.g., Coleman et al. 1966). Studies also have found that poverty, unstable families, and other social disturbances in a community are major roots of the problems in urban education (Walberg and Kopan 1972; Passow 1977; Sinclair and Ghory 1987; Council of Chief State School Officers 1988; Casserley and Kober 1990).

Unfortunately, very little systematic and comparable national data are available for researchers to further their understanding of the demographic context of inner-city children. Most information from the local school district is embedded in or mixed with the information on the larger context of urban education, which includes middle-class and affluent neighborhoods.

One reason for the lack of specific inner-city education information in the nation is the lack of a consistent definition of *inner city* and the difficulty in drawing a clear boundary of the inner city within urban areas. The so-called inner city implies disadvantaged urban communities where physical deterioration is evident and social disturbances, such as crime and illegal drugs, are widespread. These communities, however, do not have clear and consistent geographic boundaries; they may be confined to the urban center or scattered throughout the city.

This study therefore attempts to (a) develop a definition of inner-city children based on information readily available from schools and communities and (b) apply this definition to a national data source to develop a demographic and socioeconomic profile of children in the inner city. Results should further our understanding of the unique "culture" or environmental background of inner-city children and reveal special education needs of these children.

DEFINITION

This study used two criteria to identify inner-city children. The first is the community location. An inner city must be within a Standard Metropolitan Statistical Area (SMSA), a variable generally available in most national data bases. The second criterion is the community's poverty level. As mentioned earlier, an inner city implies a community where poverty and other social problems are evident. The information on these aspects is hard to collect, particularly if the information applies only to certain sections within a geographical area. A proxy measure for the socioeconomic condition is a necessary and practical alternative. One such measure available from public schools is the percentage of students participating in free or reduced-price lunch programs.

In this study, urban schools were assumed to be located in communities where a substantial number of families were on welfare if more than 50% of the students participated in free or reduced-price lunch programs. These schools were considered to be in the inner-city, and students attending them were designated as inner-city children.

For comparison, schools in suburban and rural settings were also classified by the percentage of students in free or reduced-price lunch programs. Schools in both types of communities with more than 50% of the students in free or reduced-price lunch programs were labeled as disadvantaged schools. The other schools, with less than 50% of the students in free or reduced-price lunch programs, were labeled advantaged schools.

POPULATION ESTIMATES

The preceding definition was applied to data collected by the National Educational Longitudinal Study of 1988 (NELS:88).[1] Results of this study showed that over 25% of the eighth graders (833,000 students) in 1988 were enrolled in urban schools (7% in the inner city, and another 18% in other urban settings). The total eighth-grade enrollment was estimated at 3.3 million in 1988.[2]

Assuming a similar percentage distribution of students in other grade levels from kindergarten to twelfth grade, there are over three million students in inner-city schools whose educational attainment will affect the well-being of this country.

Results of this study also show that sizable numbers of eighth graders were enrolled in suburban and rural disadvantaged schools (3% and 5%, respectively). Thus a total of 15% of all eighth graders in 1988, including children in the inner city, were attending schools in which more than 50% of students participated in free or reduced-price lunch programs.

PREDOMINANT MINORITIES

A well-known phenomenon in inner-city schools is the high concentration of racial/ethnic minorities. Based on NELS:88 data, eight of every ten inner-city eighth graders in 1988 were minorities.[3] Overall, the largest group was African American (48%), followed by Hispanic (25%). Only about 20% of students in inner-city schools were white. Smaller percentages of Asian Americans and Native Americans made up the rest (see Table 1).

This racial/ethnic distribution of students differed significantly from that of students in other communities. In advantaged suburban and rural schools, for example, over 80% of the eighth graders in 1988 were white. Only in disadvantaged suburban areas were there similar high concentrations of minorities; the majority were Hispanic.

A further examination of student distribution revealed that the composition of race/ethnicity in the inner city varied by geographic region. Although African Americans were the most dominant group in the north central region (59%), the South (50%), and the Northeast (43%); Hispanics were the most dominant group in the West (45%). As expected, there were also more Asian Americans and Native Americans in the West.

High proportions of minorities are educated in the inner city. As shown in Table 2, 25% of African American and 17% of Hispanic students were enrolled in inner-city schools, as compared to 2% of

TABLE 1

Percentage Distribution of the 1988 Eighth Graders by Race/Ethnicity in Each Type of Community

	Urban		Suburban		Rural	
Race/Ethnicity	Inner	Other	Disadvantaged	Advantaged	Disadvantaged	Advantaged
Asian	3.5	5.3	6.9	3.9	1.1	1.5
African American	48.1	18.2	15.6	8.1	21.2	7.3
Hispanic	25.4	13.2	37.0	6.8	21.3	4.1
Native American	3.0	1.0	2.2	0.8	4.9	0.9
White	20.0	62.3	38.3	80.4	51.5	86.2

white eighth graders enrolled in 1988. Thus the quality of inner-city education has a greater impact on minorities than on the majority in this country. The improvement of inner-city education is a critical step toward raising the overall education attainment level of minorities.

There is also a sizable proportion of minority students enrolled in disadvantaged schools in suburban and rural areas. These schools, together with inner-city schools, enrolled a total of 49% of Hispanic students, 40% of Native American students, and 36% of African American students. In comparison, just 7% of white students were enrolled in disadvantaged schools. Clearly, the quality of these schools affects minorities more than it affects majority students.

LANGUAGE MINORITIES

Inner-city schools are a melting pot for students of different backgrounds. About one-quarter of these students were classified as language minorities whose dominant language at home is not English. About 82% of the language minorities were Hispanic and Asian American. As expected, the percentage figures varied by region. In areas where there were more Hispanics, the percentage of language minorities was higher. Similarly, in disadvantaged suburbs where there were high concentrations of Hispanics, the percentage of language minorities was also very high (38%). In disadvantaged rural areas, it was 28% (see Table 3).

Under the Bilingual Education Act (P.L. 100-297), schools are encouraged to implement special programs using bilingual educational practices, techniques, and methods in order to provide equal educational opportunity for children with limited English proficiency and to promote educational excellence. The high concentration of language minorities in the inner city presents unique challenges to school systems. NELS:88 data showed that inner-city schools offered special language programs and have more foreign language courses for students to choose from than do other schools.

TABLE 2
Percentage Distribution of Students by Community Type within Each Race/Ethnicity Group

	Urban		*Suburban*		*Rural*	
Race/Ethnicity	*Inner*	*Other*	*Disadvantaged*	*Advantaged*	*Disadvantaged*	*Advantaged*
Asian	7.1	27.9	6.5	45.6	1.4	11.6
African American	25.0	24.8	3.8	24.3	7.3	14.8
Hispanic	17.3	23.6	11.8	27.0	9.6	10.8
Native American	16.1	14.2	5.7	26.7	17.7	19.5
White	1.9	15.8	1.7	45.1	3.3	32.2

NOTE: Totals in each row may not equal 100.0 because of rounding.

TABLE 3
Percentage of Students by Language Minority Status within Each Community Type

	Urban		*Suburban*		*Rural*	
Language Status	*Inner*	*Other*	*Disadvantaged*	*Advantaged*	*Disadvantaged*	*Advantaged*
Nonminority	75.7	85.5	62.0	91.9	72.3	94.9
Minority	24.3	14.5	38.0	8.1	27.7	5.1

UNSTABLE FAMILIES

Another widespread phenomenon in the inner city is the relatively high number of unstable families. NELS:88 data showed that less than one-half of the students in the inner city lived with both natural parents (44%), as compared to over 60% of their counterparts in other communities (see Table 4). A very high percentage of inner-city children lived with their mothers only (31%), about twice the percentage of students in other communities. Fourteen percent of inner-city children lived with their mother and a male guardian, and 6% lived with other relatives or nonrelatives.

The difference in family composition was even more pronounced among African Americans. About 42% of African American children in the inner city lived with their mother only, and less than 30% lived with both natural parents (see Table 5). High percentages of Hispanic and Native American children lived with their mother only (25% and 29%, respectively). In contrast, Asian Americans showed a high percentage of children living with both natural parents (75%).

NELS:88 data also showed that a large percentage of inner-city students' parents were unmarried (46%), including divorced, separated, never married, widowed, and cohabiting. In contrast, unmarried parents in other communities were less than 26% (see Table 6). A high percentage of children reported that their parents were never married (12%), probably indicating that many inner-city children were born out of wedlock. Consistent with data shown in Table 4, a much higher percentage of African American students' parents in the inner city were unmarried (63%), followed by Native Americans (45%), and Hispanics (36%).

Considerably more inner-city children than children in other communities live in unstable families in which children may receive inadequate care and support for success in school (Peng and Lee 1991). Such home environments are undoubtedly stressful to children and can affect teachers and school environments as well.

TABLE 4
Percentage Distribution of Students by Family Composition

	Urban		*Suburban*		*Rural*	
Family Composition	*Inner*	*Other*	*Disadvantaged*	*Advantaged*	*Disadvantaged*	*Advantaged*
Mother and father	44.0	62.0	59.7	66.9	63.5	65.6
Mother and male guardian	14.3	10.7	12.4	11.5	10.0	11.7
Father and female guardian	1.8	2.3	3.4	2.7	2.8	2.8
Mother only	30.8	19.4	17.2	13.8	17.9	14.2
Father only	2.8	2.5	3.1	2.7	1.5	2.4
Other relative/nonrelative	6.3	3.2	4.1	2.3	4.4	3.3

NOTE: Totals in each column may not equal 100.0 because of rounding.

TABLE 5
Percentage Distribution of Inner-City Students by Family Composition

Race/Ethnicity	*M and F*	*M and MG*	*F and FG*	*M only*	*F only*	*Other*
Asian	73.0	2.1	0.6	9.9	4.9	9.5
African American	29.6	16.3	1.4	41.5	2.3	8.9
Hispanic	54.3	13.6	1.8	24.6	2.5	3.5
Native American	38.1	15.8	3.3	28.5	2.5	11.8
White	60.8	12.6	3.0	18.0	3.7	2.0

NOTE: M = mother; F = father; MG = male guardian; FG = female guardian. Totals in each row may not equal 100.0 because of rounding.

UNDEREDUCATED PARENTS

Parents in inner cities and disadvantaged suburban and rural communities had significantly lower educational attainment than parents in advantaged communities. In inner cities, about 22% of parents did not graduate from high school as compared to 8% of parents in other urban communities. A similar pattern was observed in suburban and rural communities (see Table 7).

A much higher percentage of Hispanic parents (41%) and Asian American parents (38%) did not finish high school, compared to 14% of African American parents and 14% of white parents who did not finish high school (see Table 8). A high percentage of Asian American parents who did not finish high school may have been refugees from Southeast Asia.

DEPRESSING ECONOMIC CONDITIONS

The majority of children in the inner city live in poverty-stricken conditions. The unemployment rates of inner-city parents were highest among the six community types. About 15% of the inner-city mothers and about 9% of the inner-city fathers were unemployed when the parent data were collected in the spring of 1989. In the inner city, an additional 6% of mothers and 7% of fathers were either retired or disabled (see Table 9).

TABLE 6

Percentage Distribution of Students by Parents' Marital Status and Community Type

Marital Status	Urban		Suburban		Rural	
	Inner	Other	Disadvantaged	Advantaged	Disadvantaged	Advantaged
Married	52.8	74.1	74.3	80.7	75.6	79.7
Divorced	15.4	13.8	10.2	11.5	10.9	12.0
Widowed	3.8	2.7	3.6	2.4	3.1	2.0
Separated	12.7	4.3	5.9	2.6	4.1	2.8
Never married	12.0	3.1	3.9	1.2	5.1	1.6
Cohabiting	3.4	2.0	2.2	1.6	1.2	1.9

NOTE: Totals in each column may not equal 100.0 because of rounding.

TABLE 7
Percentage Distribution of Students by Parents' Highest Education Level

	Urban		*Suburban*		*Rural*	
Education Level	*Inner*	*Other*	*Disadvantaged*	*Advantaged*	*Disadvantaged*	*Advantaged*
Unknown	4.1	1.8	4.0	1.2	1.2	0.9
Did not finish high school	22.0	7.9	31.9	6.1	23.4	9.9
High school graduate or general equivalency diploma	24.0	16.2	21.2	18.4	26.2	25.1
> High school and < four-year degree	40.1	41.2	32.5	41.7	39.5	43.4
College graduate	6.3	18.0	6.1	16.7	6.4	11.8
M.A. or equivalent	2.6	8.6	2.8	10.9	2.1	6.6
Ph.D., M.D., or other	0.9	6.2	1.4	5.0	1.2	2.4

NOTE: Totals in each column may not equal 100.0 because of rounding.

TABLE 8
Percentage Distribution of Inner-City Parents' Education Level by Race/Ethnicity

Education Level	*Asian*	*Hispanic*	*African*	*White*	*Native American*
Unknown	5.4	5.3	3.4	2.0	17.2
Did not finish high school	37.8	41.2	14.3	14.2	17.0
High school graduate or general equivalency diploma	11.9	18.0	25.3	30.0	21.2
> High school and < four-year degree	30.3	27.9	48.2	38.1	40.4
College graduate	11.6	3.9	6.0	10.5	0.0
M.A. or equivalent	1.7	2.9	1.9	4.5	0.0
Ph.D., M.D., or other	1.2	0.7	0.9	0.7	4.2

NOTE: Totals in each column may not equal 100.0 because of rounding.

Family income was lowest in inner cities. As shown in Table 10, about 48% of students lived in families whose annual income was below $15,000 in 1988. In contrast, families with annual family incomes of less than $15,000 in other urban areas totaled 19% and in advantaged suburban communities only 12%. In 1988 the poverty threshold for a family of four was $12,092 and $16,149 for a family of six. The percentage of all persons in the country below the poverty threshold was about 13% (U.S. Bureau of the Census 1989).

Among racial/ethnic groups in the inner city, a higher percentage of African Americans and Hispanics had income levels below $15,000 (55% and 53%, respectively). In contrast, about 28% of white families were below this income level.

DISCUSSION

Children in the inner city differ from children in other communities in many ways. First, they are predominantly minorities. This trend is likely to intensify, because more African Americans moved from the rural South into cities (Lemann 1986a, 1986b), and more Hispanics migrated into urban areas (Wilson 1987). Minorities also

TABLE 9

Percentage Distribution of Students by Parents' Employment Status

	Urban		*Suburban*		*Rural*	
Employment Status	*Inner*	*Other*	*Disadvantaged*	*Advantaged*	*Disadvantaged*	*Advantaged*
Mother or female guardian						
Currently working	79.1	87.3	83.6	90.1	83.6	88.2
Unemployed	15.1	9.6	12.3	7.8	12.6	9.3
Retired	2.1	1.3	2.1	0.8	0.7	0.9
Disabled	3.7	1.7	4.0	1.2	3.1	1.6
Father or male guardian						
Currently working	84.3	93.0	85.0	93.7	84.1	90.9
Unemployed	8.5	3.7	7.0	3.0	7.6	3.7
Retired	3.5	1.9	3.3	1.5	3.2	2.2
Disabled	3.6	1.4	4.7	1.7	5.1	3.2

NOTE: Totals may not equal 100.0 because of rounding.

TABLE 10
Percentage Distribution of Students by Family Income Level

	Urban		*Suburban*		*Rural*	
Income Level	*Inner*	*Other*	*Disadvantaged*	*Advantaged*	*Disadvantaged*	*Advantaged*
None	1.6	0.5	1.7	0.2	0.4	0.2
< $1,000	3.3	0.6	2.6	0.4	1.8	0.6
$1,000-$2,999	4.8	1.4	4.1	0.6	3.4	1.6
$3,000-$4,999	6.2	2.1	3.2	1.1	6.1	2.2
$5,000-$7,499	9.1	2.8	6.1	2.0	5.0	3.9
$7,500-$9,999	7.5	3.4	7.6	2.1	8.2	4.3
$10,000-$14,999	15.8	7.8	16.8	5.7	16.8	10.0
$15,000-$19,999	12.2	7.6	12.2	9.1	12.8	12.4
$20,000-$24,999	11.9	9.7	13.9	9.1	11.4	12.4
$25,000-$34,999	13.2	18.5	15.1	18.4	17.9	20.8
$35,000-$49,999	10.0	20.8	11.1	24.5	10.7	20.1
$50,000-$74,999	3.4	15.3	4.7	19.6	4.3	10.4
$75,000-$99,999	0.5	4.2	0.7	5.6	0.6	2.3
$100,000 or more	0.3	5.3	0.2	5.4	0.7	1.8

NOTE: Totals in each column may not equal 100.0 because of rounding.

have higher fertility rates (Hodgkinson 1989). Second, a much greater proportion of children in the inner city come from very poor, undereducated, or unstable families. Each of these situations represents an educational disadvantage, and thus most of the children in the inner city face multiple disadvantages. To overcome these disadvantages is a tremendous challenge to students and educators. Since some communities are unable to provide the necessary resources, their education problems may be further complicated.

This study also reveals that over one-quarter of the schoolchildren in this nation were enrolled in urban schools, a significant portion of them in the inner city. The quality of schooling in the inner city and other urban areas certainly affects the educational opportunities of these children. This is particularly serious, because a large proportion of the nation's minorities are educated in inner-city schools. The effectiveness of these schools greatly affects the overall achievement of minorities. Thus the improvement of inner-city education is essential in the effort to increase the educational attainments of minorities.

The multitude of demographic disparities in the inner city points to the need for instruction that considers each students' readiness, motivation, interest, learning skills, and other factors in the prescription of classroom materials and the choice of teaching strategies. It also points to the need for educators to understand the culture of students' families — their parenting practices, value systems, attitudes toward education, as well as the health, social, and psychological problems associated with urban conditions. Educators need to know more about what strengths and deficiencies students bring to school in order to provide them with relevant and effective programs. As pointed out by Knapp and Turnbull (1990), schooling designed for children of traditional families may not fit well with the children from the inner city, since many of them do not have the kind of preparation, family support, or the mainstream "culture" typically required for success in school. Thus different schooling practices, instructional strategies, and curriculum may be required, including new ways to interact with parents to bring a closer connection among schools, families, and the community. Alternatives to conventional wisdom, as suggested by Knapp and Turnbull (1990), are worth exploring.

To be more specific, effective inner-city school programs require further understanding of the educational function taking place at home and in the community. We need to know (a) how demographic diversity and poverty have impeded the conventional or current school education, (b) how the educational needs of inner-city students differ from those of students with different socioeconomic backgrounds, (c) how the home and community environments affect students' aspirations and motivation, and (d) the extent to which parents can work with schools to bring forth the potential of their children. In other words, we need to know more about what parents do or do not do at home that affects student learning and more about how to engage parents and the community in education. Previous studies have found that parents of low socioeconomic status tend to be less involved in the schooling process, less communicative with their children, and more likely to impose strict rules at home for their children without complementary support and assistance (e.g., Peng and Lee 1991). Many inner-city families may have these problems. Thus helping parents educate their children at home is another challenge to schools and other community service organizations.

The socioeconomic disparities in inner cities also point out that many domestic problems that affect student learning cannot be resolved by schools alone. Addressing poverty and unstable families, for example, may require other social assistance. Collaboration with nonschool agencies may be needed to provide a comprehensive plan to improve education in inner cities.

NOTES

1. NELS:88 is a study sponsored by the National Center for Education Statistics, U.S. Department of Education. It involved 24,599 eighth graders from a sample of 1,035 high schools across the country selected to represent a total of about 39,000 schools with eighth graders. Within each school, approximately 26 students were randomly selected. The baseline data collection was completed in 1989. The response rate was 98.9% for schools and 93.4% for students.

The second follow-up survey is being completed, and the third follow-up survey is scheduled for 1992. Details of the study design and the data content are described in the study's data file documentation (Ingels et al. 1990). The baseline data were the bases for this study.

It should be noted, however, that the study excluded mentally handicapped students, students not proficient in English, and students having physical or emotional problems. Thus estimates of students for some subgroups are underreported.

2. The total enrollment of the seventh graders in 1987-88 in public elementary and secondary schools was 2,910,432. This number plus 7.3% of total private school enrollment in 1987-88 adds up to about 3.3 million (see National Center for Education Statistics 1990, 56, 68). Those students were assumed to be eighth graders in the 1988-89 school year when NELS:88 was conducted.

3. Because of the highly stratified and clustered sample design, all analyses in this study required the use of sample weights to obtain unbiased estimates. A sample weight is the universe of the probability of being selected, adjusted for nonresponses. Furthermore, the standard errors of statistics for this complex sample design were adjusted by a design effect. The design effect is a measure of the impact of departures from simple random sampling on the precision of sample estimates. For any statistical estimator, such as a mean or a proportion, the design effect is the ratio of the exact variance of a statistic derived from the complex sample design to that obtained from the formula for a simple random sample of the same size. Design effects for subgroups vary and are generally smaller than the design effect for the total group. For simplicity, the mean design effect of 2.5 for all students was used in this analysis. Detailed descriptions of sample weights and design effect are included in the data file *User's Manual* (Ingels et al. 1990). The following formula was used to calculate the standard error of a percentage: $SE = DEFF^{1/2} \times (p(1-p)/n)^{1/2}$, where p is the weighted percentage of respondents giving a particular response, n is the size of the sample, and $DEFF$ is the mean design effect of 2.5. All group differences cited in the text are significant at the .05 level.

REFERENCES

Casserley, M., and N. Kober. 1990. *Results 2000: Progress in meeting urban education goals.* Washington, DC: Council of the Great City Schools.

Coleman, J. S., E. Q. Campbell, C. J. Hobson, J. McPartland, A. M. Mood, F. D. Weinfeld, and R. L. York. 1966. *Equality of educational opportunity.* Washington, DC: GPO.

Council of Chief State School Officers. 1988. *School success for students at risk.* New York: Harcourt Brace Jovanovich.

Hahn, A., J. Danzberger, and B. Lefkowitz. 1987. *Dropouts in America: Enough is known for action.* Washington, DC: Institute for Educational Leadership.

Hodgkinson, H. L. 1989. *The same client: The demographics of education and service delivery systems.* Washington, DC: Institute for Educational Leadership.

Ingels, S. J., S. Y. Abraham, R. Karr, B. D. Spencer, and M. R. Frankel. 1990. *User's manual: National Educational Longitudinal Study of 1988.* Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Knapp, M. S., and B. J. Turnbull. 1990. *Better schooling for the children of poverty: Alternatives to conventional wisdom. Vol. 1, Summary.* Washington, DC: Office of Planning, Budget, and Evaluation, U.S. Department of Education.

Lemann, N. 1986a. The origins of the underclass. Part 1. *Atlantic*, June, 31-55.

Lemann, N. 1986b. The origins of the underclass. Part 2. *Atlantic*, July, 54-68.

National Center for Education Statistics. 1990. *Digest of education statistics, 1990.* Washington, DC: GPO.

Ornstein, A. C., and Levin, D. U. 1989. Social class, race and school achievement: Problems and prospects. *Journal of Teacher Education* 40(5): 17-23.

Passow, A. H. 1977. Urban education: The new challenge. *Educational Researcher* 6(9): 5-10.

Peng, S. S., and R. M. Lee. 1991. Home variables that make a difference: A study of 1988 eighth graders. Typescript.

Sinclair, R. L., and W. J. Ghory. 1987. *Reaching marginal students: A primary concern for school renewal.* Berkeley, CA: McCutchan.

U.S. Bureau of the Census. 1989. Money, income, and poverty in the United States: 1988. *Current population reports*, ser. P-60, no. 166. Washington, DC: GPO.

Walberg, H. J., and A. T. Kopan, eds. 1972. *Rethinking urban education.* San Francisco: Jossey-Bass.

Wilson, W. J. 1987. *The truly disadvantaged: The inner city, the underclass, and public policy.* Chicago: University of Chicago Press.

Educational Levels of Adolescent Childbearers at First and Second Births

DIANE SCOTT-JONES
University of Illinois at Urbana-Champaign

This study employed the 1985 birth records of a midwestern state to assess the educational levels of white, black, and Hispanic adolescents, 15–19 years of age, having first and second births. There were age and ethnic differences in the relationship of first and second births to educational levels of adolescents. Marriage was differentially related to educational level among the three ethnic groups. The educational level of fathers, when reported, was significantly correlated with the adolescent childbearers' educational level. Implications of these findings for future research and for programs and policies related to adolescent pregnancy prevention and to educational improvement are discussed.

Sexual activity among adolescents has increased in the past two decades (Centers for Disease Control 1991). In 1988, 51 percent of white and 59 percent of black adolescents were sexually active, in contrast with 27 percent of white and 46 percent of black adolescents in 1970. Birthrates of adolescent females increased in 1988, after almost two decades of decline (National Center for Health Statistics 1990). The 1988 adolescent birthrate, however, remains lower than the 1972 rate. Although the majority of American adolescents who give birth are white, the proportion of blacks among adolescent childbearers is almost double their proportion of the adolescent population. Blacks are 15 percent of the adolescent population and 29 percent of adolescent childbearers (Children's Defense Fund 1988). Hispanic adolescents have higher birthrates than do white adolescents but lower rates than blacks (Children's Defense Fund 1990).

The developmental life course of adolescents making early transitions into childbearing may be altered substantially. Other developmental transitions may affect and be affected by early childbearing. The com-

© 1991 by The University of Chicago. All rights reserved.
0195-6744/91/9904-0005$01.00

August 1991

pletion of schooling necessary for entering the adult labor force successfully is a major developmental task that adolescent childbearers must accomplish. This article reports a descriptive study of the educational levels of adolescents having first and second births.

Adolescent childbearers have lower educational attainment than do those who delay childbearing until after adolescence (Marini 1984). The negative effect of adolescent childbearing on educational attainment is greater the younger the adolescent (Mott and Marsiglio 1985). Differences in educational attainment between adolescent mothers and those who delay childbearing remain throughout the adult life course; however, educational attainment is a stronger predictor of black adult women's income than is the experience of adolescent pregnancy (Scott-Jones and Turner 1990).

Some progress in the educational attainment of adolescent mothers has been made. Currently, a greater proportion of adolescent mothers complete high school than in the past (Upchurch and McCarthy 1989). Black adolescent mothers are more likely than white adolescent mothers to complete high school (Upchurch and McCarthy 1990; McCrate 1988). Of 19-year-olds giving birth in 1988, 65 percent of black mothers and 60 percent of white mothers had completed 12 or more years of school. Of 18-year-olds giving birth in 1988, 49 percent of black mothers and 43 percent of white mothers had completed 12 or more years of school (National Center for Health Statistics 1990). An assessment of the educational attainment of black adolescent mothers 17 years after their first pregnancy indicated that two-thirds had completed high school, one-third had continued formal education beyond high school, and 5 percent were college graduates (Furstenberg et al. 1987).

Although educational attainment is depressed by the experience of adolescent pregnancy, a complicating possibility is that adolescents who are not actively engaged in school and are not performing well may be more likely to become pregnant. Young adolescents may be moving along a developmental trajectory that leads to low educational expectations and low academic achievement prior to the occurrence of pregnancy. Both black and white, and male and female, adolescents who have high educational expectations are less likely to be sexually

---

DIANE SCOTT-JONES is associate professor in the Department of Educational Psychology and the Department of Psychology at the University of Illinois at Urbana-Champaign. Her research interests include social development, family processes, the development of minorities, and social policy issues for children, youth, and families.

active than adolescents with low educational expectations (Scott-Jones and White 1990). Many pregnant adolescents had poor basic skills before the pregnancy occurred (Rindfuss et al. 1984). In the National Longitudinal Survey of Youth, adolescents who became pregnant after dropping out of high school had a greatly reduced probability of eventually graduating. For adolescent females enrolled in school, however, childbirth was not predictive of subsequently dropping out; parental education, two-parent family structure, reading materials in the home, enrollment in a college preparatory curriculum, and not smoking or drinking were predictive of high school graduation (Upchurch and McCarthy 1990). In Project Redirection, an intervention program for pregnant and parenting adolescents, of those who had dropped out of school, one-half had done so before the pregnancy occurred (Polit et al. 1988).

Once a pregnancy occurs during adolescence, the adolescent may be diverted into a life course that emphasizes childbearing and child rearing, to the exclusion of continued schooling. The adolescent mother's subsequent pattern of childbearing may affect educational attainment. Although the difference in completed family size between adolescent and older childbearers has declined, women who have their first pregnancy during adolescence have more children than do women who have their first pregnancy after the adolescent years (Teachman 1985). In 1985, 1 percent of 15–19-year-old females had a repeat birth (Children's Defense Fund 1988). For a majority of participants in an intervention program focusing specifically on the prevention of a second pregnancy during adolescence, a second pregnancy occurred within two years of the first (Polit and Kahn 1986).

Adolescent childbearers currently are less likely to be married than were adolescent childbearers in the past (National Center for Health Statistics 1990). Adolescents in general are unlikely to be married, and there are racial differences in marriage rates. In 1985, of all 15–19-year-olds, 11.2 percent of Hispanics, 7.6 percent of whites, and 1.6 percent of blacks were married (Children's Defense Fund 1988). Almost two-thirds of white adolescents and almost all black adolescents having first births become pregnant outside marriage (Furstenberg et al. 1989). The role of marriage in educational outcomes for adolescent mothers may not be positive, however. Married black and Hispanic adolescent mothers are not likely to remain in school; further, their divorce rate is high (Children's Defense Fund 1990; McLaughlin et al. 1986). Similarly, black and white males who marry in adolescence have lower educational attainment and more marital disruption than do males who marry in adulthood; the relative educational disadvantage of having married in adolescence remains throughout the adult life course

(Teti et al. 1987). The presence or timing of marriage was not associated with educational attainment of adult black and white women who had their first birth in adolescence (Teti and Lamb 1989).

Whether marriage occurs or not, adolescent mothers' continuing their schooling may be influenced by the level of education their male partners have attained. Adolescent fathers have higher dropout rates than their peers who have not fathered children; dropout rates are higher for married adolescent fathers than for those who are unmarried (Marsiglio 1987). The fathers of children born to adolescent mothers are not always themselves adolescents, however. In 1985, 37 percent of adolescent mothers who gave birth did not identify the father. Only 18 percent of the fathers were reported to be 15–19 years of age; the remaining fathers were 20 years or older (National Center for Health Statistics 1987).

This study employs the 1985 birth records of a midwestern state, Illinois. In 1985, Illinois ranked fifth among the states in the nation in the numbers of births to adolescents; however, the birthrate of 15–19-year-olds in Illinois in 1985, 49.3 births per 1,000 females, is not substantially different from the 1985 national rate of 51.2 births per 1,000 females (Children's Defense Fund 1988).

Illinois birth records were used to answer the following questions about the educational levels of adolescent childbearers. (1) Does the educational level of adolescents experiencing a second birth differ significantly from those experiencing a first birth, among white, black, and Hispanic adolescents of different ages? If adolescent pregnancy results in a decline in educational level, then adolescents experiencing a second birth should have a lower educational level than that of first-time childbearers. (2) Is the educational level of adolescent childbearers, at the time of birth, significantly below that expected for their age? If low educational achievement is an antecedent of early childbearing, then educational level should be below age norms at the time of a first birth, as well as at the time of a second birth. (3) Is educational level different for married and unmarried childbearers? Married adolescents were expected to have lower educational levels than unmarried adolescents. (4) Among second-time childbearers, does the outcome of the first pregnancy affect educational level? It was expected that second-time childbearers whose pregnancy resulted in live birth would have lower educational levels than those whose first pregnancy did not result in a live birth. (5) Are age and educational level of fathers related to mothers' educational levels? It was expected that older and less educated fathers would be associated with lower educational levels for adolescent childbearers.

Method

Data Source

Data used are 1985 birth records for the state of Illinois from the Illinois Department of Public Health, Vital Records. In Illinois in 1985, 179,004 births were recorded; mothers ranged in age from 11 years to 48 years. Of these births, 12.7 percent were to adolescents 19 years of age or younger; .3 percent were to adolescents younger than 15 years of age. For blacks, 26.8 percent of all births were to adolescents, for Hispanics, 14.9 percent, and for whites, 8.9 percent.

Of first births in 1985, 26.4 percent were to adolescents 19 years of age or younger. For blacks, however, 54.5 percent of first births were to adolescents, for Hispanics, 36.2 percent, and for whites, 18.1 percent. For first births in 1985, the mean age of mothers was 24.5 years for whites, 20.2 years for blacks, and 21.8 years for Hispanics.

In 1985, 22,667 births were to adolescents. Of these, two-thirds (15,276) were first births. Approximately one-fourth (5,501) were second births. Third or later births were 8 percent (1,890) of births to adolescents. Table 1 shows the distribution of first and second births to adolescents, by age and ethnic group. (Number of cases in the analyses reported in the Results section may differ slightly from those in table 1 because of missing data.) Blacks have a disproportionately high percentage of first births (40 percent) and second births (51 percent). Both first and second births to adolescents increase with age within all ethnic groups. The age distribution of adolescent births differs among the three ethnic groups, however. Younger adolescents account for a higher percentage of births to blacks than to the other two groups; younger adolescents account for a higher percentage of births to Hispanics than to whites.

The marital status of adolescents at first and second births, by age and ethnic group, is presented in table 1. The percentage of adolescents who are married increases with age within each ethnic group. By the age of 19 years, more than half of white and Hispanic adolescents having first or second births are married. For blacks, however, the percentage married ranges from practically none at the age of 15 years to less than 10 percent at the age of 19 years.

Because there is no measure of socioeconomic status, race/ethnic origin of mothers may be confounded with socioeconomic status in the analyses. Further, in this data set, race/ethnic origin of adolescent mothers is confounded with rural-urban residence. The majority of

TABLE 1

First and Second Births, by Age, Ethnicity, and Marital Status

	WHITE				BLACK				HISPANIC			
AGE	First Birth (*n* = 7,550)	Percentage Married	Second Birth (*n* = 2,138)	Percentage Married	First Birth (*n* = 6,028)	Percentage Married	Second Birth (*n* = 2,788)	Percentage Married	First Birth (*n* = 1,698)	Percentage Married	Second Birth (*n* = 575)	Percentage Married
<15	102	. . .	3	. . .	418	. . .	20	. . .	39	. . .	2	. . .
15	334	9.0	15	20.0	707	.4	101	0	99	16.2	9	44.4
16	840	29.0	86	37.2	1,131	.7	315	2.2	238	22.3	37	21.6
17	1,512	39.0	281	38.4	1,259	3.0	570	1.9	378	39.4	97	36.1
18	2,158	50.1	662	54.4	1,324	4.9	830	6.4	446	46.4	186	48.9
19	2,604	61.4	1,091	64.5	1,189	8.2	952	9.7	498	60.4	244	53.3

births to white adolescents, 79 percent of first births and 81 percent of second births, were to those residing in small cities and townships. The majority of births to black and Hispanic adolescents were to those residing in Chicago, the only major urban center in the state, with a population exceeding 3 million. For blacks, 70 percent of first births and 69 percent of second births were to adolescents residing in Chicago; for Hispanics, 77 percent of first births and 73 percent of second births were to adolescents residing in Chicago.

Variables employed in this study included race/ethnic origin of mother, age of mother, age of father, educational level of mother, educational level of father, mother's marital status, and total number of children born to the mother. Most of the data are self-reported and are subject to the usual limitations of self-report data. Birth data, however, are generally considered highly reliable (Hayes 1987). These inclusive birth records are important, particularly for examining whether educational level is low prior to the occurrence of a first pregnancy. Because of the small number of births to adolescents younger than 15 years, the major analyses of this study employed the birth records of 15–19-year-old adolescents.

Results

First and Second Births

Analyses of variance were conducted to determine whether educational level differed for adolescents experiencing a second birth and those experiencing a first birth in the three ethnic groups. Separate two (birth: first, second) × three (ethnic group: white, black, Hispanic) analyses of variance were conducted for each of the five age groups from 15 to 19 years. Analyses were conducted separately by age because the maximum possible years of schooling for adolescents to have completed varies within this age range. Births to adolescents younger than 15 years were not included because of the small numbers. Mean years of education at first and second births, by age and ethnic group, are presented in table 2.

For 15-year-olds, there were no significant main effects of first or second birth and no significant interaction. There was a significant main effect of ethnic group (F (2, 1,254) = 14.62, $P < .0001$). Tukey's studentized range (HSD) test indicated that Hispanics had significantly lower educational levels than did blacks and whites.

August 1991

TABLE 2

Mean Years of Education of White, Black, and Hispanic Adolescents Having First and Second Births

	WHITE			BLACK			HISPANIC		
AGE	National Median	First Birth	Second Birth	National Median	First Birth	Second Birth	National Median	First Birth	Second Birth
15	9.1	9.1	9.0	8.8	9.1	9.0	8.7	8.6	8.2
16	10.1	9.9	9.6	9.7	10.0	9.9	9.7	9.1	8.0
17	11.1	10.7	10.0	10.9	10.8	10.7	10.5	9.4	8.8
18	12.1	11.4	10.8	11.8	11.5	11.3	11.3	9.7	8.8
19	12.6	11.8	11.1	12.5	11.9	11.6	12.1	9.9	9.3

NOTE.—The national medians reported are median school years completed for these ages from the U.S. Bureau of the Census (1987). For whites and blacks the medians for years of education are for 1985, for females only. For Hispanics, data were not available for females only; the medians reported for Hispanics include both males and females.

For each of the remaining age groups, there were significant first and second pregnancy × ethnic group interactions; for 16-year-olds, F (2, 2,631) = 11.9, $P < .0001$; for 17-year-olds, F (2, 4,083) = 15.4, $P < .0001$; for 18-year-olds, F (2, 5,575) = 22.3, $P < .0001$; and for 19-year-olds, F (2, 6,536) = 8.43, $P < .0002$.

Post hoc pairwise comparisons of means were conducted. All effects reported were significant at $P < .0001$. For 16-year-olds, Hispanics had significantly lower educational levels than did whites or blacks; adolescents having second births had significantly lower educational levels than did those having a first birth among Hispanics only. For the remaining age groups, Hispanics having a first birth had significantly lower educational levels than did whites or blacks having first or second births; numbers of Hispanics having second births were significantly lower than Hispanics having first births.

For both 17- and 18-year-olds, blacks having a first or second birth and whites having a first birth were not significantly different from one another and had significantly higher educational levels than whites having a second birth. For 19-year-olds, blacks and whites having a first birth were not significantly different from one another and had significantly higher educational levels than did the remaining groups. Blacks having a second birth had significantly higher educational levels than did whites having a second birth.

Comparisons to National Median Educational Levels

The years of education of adolescents having a first birth and those having a second birth were compared to national norms for educational attainment for white and black female adolescents in 1985 from the U.S. Bureau of the Census (1987). For Hispanics, data were not available for females only; therefore, the educational attainment of Hispanic adolescent childbearers was compared to norms for Hispanic females and males combined. Males are more likely than females to have been retained in a grade (Dryfoos 1990); therefore, the comparisons for Hispanic childbearers may underestimate their difference from Hispanic females in general.

To test whether the educational level of adolescents having a first birth and those having a second birth differed significantly from the national median for the age-ethnic group, difference scores were calculated. The educational level of the adolescent childbearer was subtracted from the national median for her age-ethnic group. Separate t-tests were conducted to determine whether the difference scores were

Adolescent Childbearers

TABLE 3

Difference between Median National Educational Levels and Educational Levels of Adolescent Mothers

	WHITE		BLACK		HISPANIC	
AGE	First Birth	Second Birth	First Birth	Second Birth	First Birth	Second Birth
15	NS	NS	.26***	.22**	NS	NS
16	−.21***	−.53***	.26***	.19***	−.57***	−1.67***
17	−.41***	−1.05***	−.11***	−.24***	−1.02***	−1.71***
18	−.72***	−1.33***	−.35***	−.55***	−1.57***	−2.54***
19	−.80***	−1.48***	−.56***	−.93***	−2.20***	−2.79***

** $P < .01$.
*** $P < .001$.

significantly different from 0. The mean difference scores are presented in table 3.

For 15-year-old adolescents having first and second births, educational level was not significantly lower than the national median; for black adolescents, educational level was significantly higher than the national median, although the magnitude of the difference was small. For 16-year-old black adolescents having first and second births, educational level was significantly higher than the national median, although, again, the magnitude of the difference was small. For both first-time and second-time childbearers, in all other age-ethnic groups, educational level was significantly lower than the national median.

Marital Status

Mean years of education for married and unmarried adolescents at first and second births, by age and ethnic group, are presented in table 4. Births to adolescents younger than 18 years were not included because of the small numbers of married adolescents, especially among blacks. To determine whether educational level differed for married and unmarried adolescents, a three (ethnic group) × two (marital status) × two (birth: first or second) × two (age: 18 or 19 years) analysis of variance was conducted. Significant interactions were found for marital status and ethnic group (F (2, 12,113) = 96.9, $P < .0001$), marital status and first and second births (F (1, 12,114) = 29.1, $P < .0001$), and ethnic group and first and second births (F (2, 12,113) = 15.68, $P < .0001$).

TABLE 4

Mean Years of Education of Married and Unmarried 18- and 19-Year-Old Mothers Having First and Second Births

	WHITE		BLACK		HISPANIC	
NUMBER OF BIRTHS AND AGE	Married	Unmarried	Married	Unmarried	Married	Unmarried
First births:						
18	11.5	11.3	11.8	11.4	9.3	10.1
19	11.9	11.7	12.3	11.9	9.5	10.5
Second births:						
18	10.8	10.7	11.5	11.2	8.2	9.3
19	11.1	11.1	11.8	11.6	9.8	8.8

Adolescent Childbearers

Post hoc pairwise comparisons of means were conducted. All significant effects reported below were significant at $P < .0001$. For 18- and 19-year-olds having first and second births, married Hispanic adolescents had significantly lower educational levels than all other groups; unmarried Hispanics were significantly lower than black and white married and unmarried adolescents. Among Hispanics, marriage was associated with the loss of approximately one year of education. For first births at 18 and at 19 years, married blacks had significantly higher educational levels than did all other groups; unmarried blacks and unmarried and married whites were not significantly different from one another. For second births at 18 and at 19 years, married and unmarried blacks did not differ significantly from each other and had significantly higher educational levels than did married and unmarried whites, who were not different from each other.

First-Birth Outcomes and Second Births

To determine whether the educational level of adolescents having a second birth varied according to the outcome of the first birth, comparisons were made between those whose first pregnancy resulted in a live birth and those whose first pregnancy did not. Adoption is rarely chosen by adolescent childbearers (Dryfoos 1990); therefore, one can assume that adoption was rarely the outcome for adolescents whose first pregnancy resulted in a live birth. Adolescents having a second birth were omitted from this analysis if their first pregnancy resulted in a live birth but the child was not living at the time of the second birth; 1.4 percent of the second births were omitted for this reason.

Among white adolescents having a second birth, the first pregnancy did not result in a live birth for 67 percent of 15-year-olds, 58 percent of 16-year-olds, 46 percent of 17-year-olds, 37 percent of 18-year-olds, and 32 percent of 19-year-olds. Among blacks, the percentages were 32 percent for 15- and 16-year-olds, 26 percent for 17-year-olds, 24 percent for 18-year-olds, and 22 percent for 19-year-olds. For Hispanics, the percentages were 18 percent for 15-year-olds, 27 percent for 16-year-olds, 17 percent for 17-year-olds, 23 percent for 18-year-olds, and 19 percent for 19-year-olds.

Separate two (outcome of first pregnancy) × three (ethnic group) analyses of variance were conducted for each of the five age groups from 15 through 19 years. For 15-year-olds, there were no significant effects; however, the sample size was small ($n = 119$). For 16-year-olds, there was a main effect of ethnic group ($F(2, 428) = 41.4$, $P < .0001$). Tukey's studentized range (HSD) test indicated that Hispanics were significantly lower than blacks and whites; the latter two groups were not significantly different from one another.

Significant first-pregnancy outcome × ethnic group interactions occurred in the separate ANOVAs for 17-year-olds (F (2, 931) = 8.0, $P < .0004$), 18-year-olds (F (2, 1,647) = 5.49, $P < .004$), and 19-year-olds (F (2, 2,243) = 2.9, $P < .05$). Pairwise comparisons of means following the separate ANOVAs for 17-, 18-, and 19-year-olds indicated that, for both whites and Hispanics, those whose first pregnancy resulted in a live birth had significantly lower educational levels than those whose first pregnancy did not result in a live birth; educational level was not significantly different for blacks whose first pregnancy resulted in a live birth and blacks whose first pregnancy did not result in a live birth. The educational level of blacks was significantly higher than that of every other group except for whites whose first pregnancy did not result in a live birth.

Educational Levels of Fathers

Of all births in 1985, 12.7 percent were to females 19 years of age or younger, but only 4.8 percent of fathers of all births were 19 years of age or younger. Of the adolescent fathers, 74 percent were 18 or 19 years of age. Table 5 presents the mean age and educational level of fathers of births to adolescent females, by mother's age and mother's ethnicity. Father's age and educational level were not reported by some adolescent mothers. The percentage of 15–19-year-old mothers in the three ethnic groups reporting father's age ranged from 79 percent to 85 percent; the percentage reporting father's educational level ranged from 54 percent to 73 percent.

The reported ages of fathers ranged from 12 years to 64 years. At each age level, the mean age of fathers reported for Hispanic adolescents was approximately one year greater than father's age reported for white adolescents, which was slightly higher than that reported for black adolescents. At each age level, the mean age of fathers was approximately four years greater than mother's age for Hispanics and approximately three years greater for whites and blacks. Father's educational level reported for Hispanics is lower than that reported for whites and blacks, and the difference is greater for older than for younger adolescents.

Pearson product-moment correlations indicated that father's educational level is positively correlated with mother's educational level within the three ethnic groups for all age groups combined (for whites, $r = .36$, $P < .0001$; for blacks, $r = .39$, $P < .0001$; for Hispanics, $r =$

TABLE 5

Mean Age and Education of Fathers of First and Second Births to White, Black, and Hispanic Adolescents

	WHITE		BLACK		HISPANIC	
MOTHER'S AGE	Father's Age	Father's Education	Father's Age	Father's Education	Father's Age	Father's Education
15	18.8	10.9	18.0	10.8	19.6	9.5
16	19.4	11.2	18.9	11.3	20.4	10.0
17	20.2	11.4	19.9	11.6	21.3	9.7
18	21.2	11.7	20.9	11.8	22.1	9.7
19	22.2	11.9	22.0	12.0	23.1	9.4

.57, $P < .0001$). The relationship of father's age to mother's educational level varies for the three ethnic groups (for whites, $r = 0$; for blacks, $r = .17$, $P < .0001$; for Hispanics, $r = -.15$, $P < .0001$).

Discussion

This study found age and ethnic differences in the relationship of first and second births to educational levels of adolescents. Marriage was differentially related to educational level among the three ethnic groups. The educational level of fathers, when reported, was significantly correlated with the adolescent childbearer's educational level. These findings have implications for future research and for programs and policies related to adolescent pregnancy prevention and to educational improvement.

Young adolescent childbearers were not as far behind the national median for their age-ethnic group as were older adolescent childbearers. An educational trajectory different from that of their age-ethnic group was not evident for 15-year-olds having a first or second birth in any ethnic group. Further, at 15 and 16 years of age, black adolescents having first and second births showed a small but statistically significant increase over the national medians. These findings suggest that interventions to maintain adolescents' engagement in school need to be instituted in early adolescence, when, even with a first or second birth, adolescents are not behind national medians for educational attainment for their age. The requirement of compulsory school attendance until age 16 years may play a role in young adolescent childbearers' remaining in school. Additional policies and programs, such as child-care programs requiring the adolescent mother to remain in school, may boost the educational attainment of older adolescent childbearers.

The finding that the youngest adolescent childbearers were not below national medians in educational attainment also suggests that young adolescents generally have trouble educationally, for reasons other than the occurrence of a first or second birth. The educational level of young adolescents, particularly blacks and Hispanics, is cause for concern, without the occurrence of pregnancy. In general, black and Hispanic students have a higher probability of being retained in grade than do white students (Dryfoos 1990). Of eighth graders in the National Education Longitudinal Study, 18 percent reported having repeated at least one grade. In that study, approximately 26 percent of black and 23 percent of Hispanic eighth graders reported having repeated at least one grade, in contrast to approximately 16 percent of whites (Hafner et al. 1990). Preventing educational failure is needed for young adolescent childbearers and for young adolescents generally.

Difference from the national medians became more pronounced for older adolescent childbearers. Adolescents having a first child at the age of 19 years could have completed high school prior to the pregnancy, if they had progressed through school on schedule. In all three ethnic groups, however, the average educational level for 19-year-olds having a first birth was less than 12 years. Hispanic adolescents having a first birth at the age of 19 years were more than two years behind the national median for 19-year-old Hispanics. These adolescents had completed, on average, less than 10 years of schooling. These findings strongly suggest that educational difficulties occurred prior to the experience of pregnancy.

The difference in educational level between adolescents having a first birth and those having a second birth varied with age and ethnic group. Among 15-year-olds and, with the exception of Hispanics, 16-year-olds, adolescents having a second birth did not have significantly lower educational levels than those having a first birth. For 17- and 18-year-olds, second births were associated with lower educational levels relative to first births for Hispanics and whites. For 19-year-olds, second births were associated with lower educational levels relative to first births for Hispanics, whites, and blacks. Black adolescents, who accounted for more than one-half of second births, were least affected educationally by second births.

The sharp ethnic differences in this study are consistent with existing literature. Hispanic childbearers had significantly lower educational levels than blacks or whites, within each age level of 15–19-year-olds. Black adolescent childbearers deviated less from the national median educational level for their age, ethnic, and gender group than did whites or Hispanics. One possible explanation involves the differential rates of childbearing among blacks, whites, and Hispanics nationally. A higher proportion of black adolescents become childbearers than do whites or Hispanics; therefore, a higher proportion of childbearers would be included in the calculation of the national median educational level for blacks than for whites or Hispanics. This could not account entirely for the findings for blacks, however. Nationally, a larger proportion of older black adolescents become childbearers than do younger black adolescents; yet older black adolescents in this study differ more from national medians for their age groups than do the younger black adolescents.

Another possible explanation for adolescent mothers' educational progress is the availability of special programs. Both black and Hispanic adolescent mothers, however, were predominantly residents of Chicago, a large urban area where one would expect to find many services for adolescent childbearers. Yet, the black and Hispanic childbearers were different in educational attainment. Another possibility is that available programs may be targeted toward blacks more than toward other ethnic groups; alternatively, programs may come to be identified informally by community members as "black" programs and may not be used extensively by other groups who need services. Another possible explanation is that black adolescent childbearers receive more support in a variety of ways in families and communities.

Whatever the reason for black adolescent childbearers' seeming resilience, their educational level is even more striking when one considers the overall educational difficulties of black and Hispanic students. The educational level of the black adolescent childbearers is not satisfactory, however. The mean educational level for 19-year-old first-time black childbearers was less than graduation from high school, which is unsatisfactory given that even graduation from high school is not likely to improve substantially the economic conditions of adult women (Moore and Wertheimer 1984).

Among Hispanics and whites, the percentage of childbearers who were married increased with age; approximately one-half of 18- and 19-year-old childbearers in these two ethnic groups were married. Although married childbearers also increased with age for blacks, marriage occurred infrequently for older black adolescent childbearers and hardly occurred at all for younger black adolescent childbearers. For 18- and 19-year-olds, marriage was associated with lower educational levels for Hispanics, was associated with higher educational levels for blacks having first births, and was not associated with educational level for blacks having second births or for whites. Thus, although blacks were much less likely to be married than were whites or Hispanics, marriage was associated with increased schooling for black adolescents at first birth.

The role of marriage in the developmental life course of adolescent childbearers is especially problematic. American society places a high value on advanced education and on marriage. Adolescent childbearers may not be able to combine school attendance and marriage successfully. Black adolescent childbearers appear to manage schooling relatively more successfully than they do marriage; Hispanic adolescent childbearers are more likely than blacks to marry but not as likely to manage school well.

The outcomes for Hispanic childbearers, with lower educational levels at first and second births and further depression of educational levels with marriage, should be studied further. Research should focus on specific subgroups labeled Hispanic. Data from the National Center

August 1991

for Health Statistics indicate that, in 1985, births to Hispanic adolescents nationally were not evenly distributed across various subgroups of Hispanics. Of births to Hispanic adolescents, 69 percent were to Mexicans, 12 percent were to Puerto Ricans, and 1 percent to Cubans (Children's Defense Fund 1988).

A relatively high percentage of adolescent childbearers provided information about the father. Fathers of births to Hispanics tended to be older and less educated than did fathers of births to blacks or whites. As expected, the majority of fathers were older than 19 years; the average age of fathers of births to blacks and whites was three years greater than mother's age and, to Hispanics, four years. The educational level of fathers, when it was reported, showed a surprisingly strong correlation with the educational level of the mothers. Because these fathers are older, it may be difficult to reach them directly with typical school programs. School-based programs will thus miss a major actor in the scenarios that lead to unplanned early pregnancy. Further, the marriage prospects of adolescent childbearers may be tied to the status of males in early adulthood. Bowman (1990) has found that a high proportion of black young adult males are jobless and experience a phenomenon called "job-search discouragement," which affects young males' capacity to provide material and emotional support for a family. More research is needed on the male partners involved in adolescent pregnancy. As Parke and Neville (1987) point out, we know little about the patterns of involvement between male partners and adolescent mothers and little about differences between adolescent and older male partners.

These results suggest the importance of studying the changing patterns of childbearing in white, Hispanic, and black adolescents' lives. Changing patterns of childbearing should be studied in relation to other transitions adolescents make, especially regarding the completion of schooling and marriage. Programs to increase educational achievement and programs to prevent unplanned pregnancies indirectly by focusing on educational and other life options need careful evaluations. Evaluations should identify the effective components of such programs and determine whether such programs are differentially used by or are differentially successful with various ethnic and age groups. Programs that focus on life options, however, have not yet produced rigorous evaluations (Hofferth 1987), and programs that focus generally on improving educational outcomes have not assessed adolescent pregnancy as part of their evaluations (Hayes 1987). Intervening in early adolescence, when childbearing is relatively infrequent, will probably be maximally effective in preventing unplanned pregnancies and in fostering the educational progress of adolescents who become pregnant.

Note

The source of data is Illinois Department of Public Health, Vital Records, 1980–88 sterilized birth tapes; the data analyses and interpretations, however, are the author's and do not reflect the official position of the Illinois Department of Public Health. The work reported herein was supported under the Educational Research and Development Center Program (agreement no. R117Q00031) as administered by the Office of Educational Research and Improvement, U.S. Department of Education, in cooperation with the U.S. Department of Health and Human Services. The findings and opinions expressed in this article do not reflect the position or policies of the Office of Educational Research and Improvement, the U.S. Department of Education, or the U.S. Department of Health and Human Services. I acknowledge the assistance of Mary Zwover Anderson in data analyses. The author thanks Joyce Epstein, Anne C. Peterson, and three anonymous reviewers for their helpful comments on this article.

References

Bowman, P. J. "The Adolescent-to-Young Adult Transition: Discouragement among Jobless Black Youth." *New Directions for Child Development* 46 (Winter 1990): 87–105.

Centers for Disease Control. "Premarital Sexual Experience among Adolescent Women—United States, 1970–1988." *Morbidity and Mortality Weekly Report* 39 (January 4, 1991): 929–32.

Children's Defense Fund. *Teenage Pregnancy: An Advocate's Guide to the Numbers*. Washington, D.C.: Children's Defense Fund, 1988.

Children's Defense Fund. *Latino Youths at a Crossroads*. Washington, D.C.: Children's Defense Fund, 1990.

Dryfoos, J. G. *Adolescents at Risk: Prevalence and Prevention*. New York: Oxford University Press, 1990.

Furstenberg, F. F., J. Brooks-Gunn, and L. Chase-Lansdale. "Teenaged Pregnancy and Childbearing." *American Psychologist* 44 (February 1989): 313–20.

Furstenberg, F. F., J. Brooks-Gunn, and S. P. Morgan. *Adolescent Mothers in Later Life*. Cambridge: Cambridge University Press, 1987.

Hafner, A., S. Ingels, B. Schneider, and D. Stevenson. *A Profile of the American Eighth Grader: NELS:88 Student Descriptive Summary*. Washington, D.C.: Government Printing Office, 1990.

Hayes, C. D. "Adolescent Pregnancy and Childbearing: An Emerging Research Focus." In *Risking the Future: Adolescent Sexuality, Pregnancy, and Childbearing*, vol. 2, edited by S. L. Hofferth and C. D. Hayes. Washington, D.C.: National Academy, 1987.

Hofferth, S. L. "The Effects of Programs and Policies on Adolescent Pregnancy and Childbearing." In *Risking the Future: Adolescent Sexuality, Pregnancy, and Childbearing*, vol. 2, edited by S. L. Hofferth and C. D. Hayes. Washington, D.C.: National Academy, 1987.

McCrate, E. "Teenage Mothers: Poverty's Vicious Cycle." *UCLA Center for Afro-American Studies Newsletter* 11 (1988): 6–7.

August 1991

Adolescent Childbearers

McLaughlin, S. D., W. R. Grady, J. G. Billy, N. S. Lansdale, and L. D. Winges. "The Effects of the Sequencing of Marriage and First Birth during Adolescence." *Family Planning Perspectives* 18 (1986): 12–18.

Marini, M. "Women's Educational Attainment and the Timing of Entry into Parenthood." *American Sociological Review* 49 (1984): 491–511.

Marsiglio, W. "Adolescent Fathers in the United States: Their Initial Living Arrangements, Marital Experience, and Educational Outcomes." *Family Planning Perspectives* 19 (1987): 240–51.

Moore, K. A., and R. F. Wertheimer. "Teenage Childbearing and Welfare: Preventive and Ameliorative Strategies." *Family Planning Perspectives* 16 (1984): 285–89.

Mott, F. L., and W. Marsiglio. "Early Childbearing and Completion of High School." *Family Planning Perspectives* 17 (1985): 234–37.

National Center for Health Statistics. *Advance Report of Final Natality Statistics, 1985*, vol. 36, no. 4, supplement. Washington, D.C.: Department of Health and Human Services, 1987.

National Center for Health Statistics. *Advance Report of Final Natality Statistics*, vol. 39, no. 4, supplement. Washington, D.C.: Department of Health and Human Services, 1990.

Parke, R. D., and B. Neville. "Teenage Fatherhood." In *Risking the Future: Adolescent Sexuality, Pregnancy, and Childbearing*, vol. 2, edited by S. L. Hofferth and C. D. Hayes. Washington, D.C.: National Academy, 1987.

Polit, D. F., and J. R. Kahn. "Early Subsequent Pregnancy among Economically Disadvantaged Teenage Mothers." *American Journal of Public Health* 76 (1986): 167–71.

Polit, D. F., J. C. Quint, and J. A. Riccio. *The Challenge of Serving Teenage Mothers: Lessons from Project Redirection*. New York: Manpower Demonstration Research, 1988.

Rindfuss, R. R., C. St. John, and L. L. Bumpass. "Education and the Timing of Motherhood: Disentangling Causation." *Journal of Marriage and the Family* 46 (November 1984): 981–84.

Scott-Jones, D., and S. L. Turner. "The Impact of Adolescent Childbearing on Educational Attainment and Income of Black Females." *Youth and Society* 22 (1990): 35–53.

Scott-Jones, D., and A. B. White. "Correlates of Sexual Activity in Early Adolescence." *Journal of Early Adolescence* 10 (May 1990): 221–38.

Teachman, J. D. "The Declining Significance of First-birth Timing." *Demography* 22 (1985): 185–98.

Teti, D. M., and M. E. Lamb. "Socioeconomic and Marital Outcomes of Adolescent Marriage, Adolescent Childbirth, and Their Co-occurrence." *Journal of Marriage and the Family* 51 (1989): 203–12.

Teti, D. M., M. E. Lamb, and A. B. Elster. "Long-Range Socioeconomic and Marital Consequences of Adolescent Marriage in Three Cohorts of Adult Males." *Journal of Marriage and the Family* 49 (1987): 499–506.

Upchurch, D., and J. McCarthy. "Adolescent Childbearing and High School Completion in the 1980's: Have Things Changed?" *Family Planning Perspectives* 21 (1989): 199–202.

Upchurch, D., and J. McCarthy. "The Timing of a First Birth and High School Completion." *American Sociological Review* 55 (1990): 224–34.

U.S. Bureau of the Census. "Educational Attainment in the United States: March 1982 to 1985." *Current Population Reports*, ser. P-20, no. 415. Washington, D.C.: Government Printing Office, 1987.

Journal of Educational Psychology
1991, Vol. 83, No. 2, 201-211

Copyright 1991 by the American Psychological Association, Inc.
0022-0663/91/$3.00

Explaining Within-Semester Changes in Student Effort in Junior High School and Senior High School Courses

Douglas J. Mac Iver
Center for Research on Effective Schooling for Disadvantaged Students, Johns Hopkins University

Deborah J. Stipek and Denise H. Daniels
Graduate School of Education, University of California, Los Angeles

Within any course, as a semester progresses some students reduce their effort and others try harder. Virtually every cognitive theory of motivation suggests that changes in ability perceptions partially determine these changes in effort. Researchers have also cited changes in students' valuing of the course and changes in extrinsic pressures as determinants of effort changes. Covariance structure modeling was used to test 4 alternative models concerning the determinants of effort in a sample of 167 junior high school and 155 senior high school students. Models specifying a direct effect of ability-perception change on effort change fit the data better than did models specifying only indirect effects or no effect of ability perceptions on effort. Ability-perception changes also directly affected students' valuing of the subject matter. The results emphasize the importance of helping students develop confidence in their abilities.

Within almost any course, as a semester progresses some students reduce their effort and others try harder. What factors are responsible for within-semester changes in student effort? In this study, we test theoretical models concerning the determinants of student effort during junior high school and senior high school.

During the past two decades, ability perceptions have come to play a central role in many theories of human motivation and action (e.g., Bandura, 1977; Covington & Beery, 1976; Dweck, 1986; Kukla, 1978; Meyer, 1987; Nicholls, 1984; Raynor & Brown, 1985). For example, according to self-efficacy theory, individuals with self-percepts of low ability are easily discouraged by failure to attain the standards they set for themselves, whereas those who are confident of their ability typically intensify their efforts when failure occurs and persist until they succeed (Bandura & Cervone, 1983). Self-worth theorists (Covington & Beery, 1976) have claimed that students who lose confidence in their ability may adopt counterproductive, effort-avoidant strategies so that failure, if it occurs, can be blamed on insufficient effort rather than on low competence. According to attribution theory, attributing failure to low ability is among the causes of learned helplessness and depression (Abramson, Seligman, & Teasdale, 1978; Weiner, 1986).

The emergence of self-perceptions of ability as a cornerstone of theories dealing with achievement-related behavior is not surprising given the importance of the expectancy construct in the pioneering theoretical work of Atkinson (1957). Atkinson defined expectancy for success at a task in terms of perceptions of the probability of success. He and his colleagues (e.g., Atkinson & Birch, 1978) have typically operationalized *subjective expectancy for success* as "perceived task facility" rather than as "perceived self-concept of ability" (Reuman, 1986). That is, Atkinson emphasized the role of easy tasks (rather than the role of high-ability perceptions) in producing high expectancies. Consequently, the most frequently adopted strategy for manipulating expectancies in classic experimental research on persistence and choice (e.g., Feather, 1961) has been to supply subjects with information about the normative difficulty of the task they are attempting. Recent reformulators of Atkinson's achievement motivation theory (e.g., Raynor & Brown, 1985; Reuman, 1986) have pointed out that high-ability perceptions lead to higher expectancies for success than do low-ability perceptions. Recent formulations, therefore—though not disregarding the role of task facility in influencing expectancies—have emphasized the role of personal ability perceptions.

The positive role of ability perceptions in influencing effort, especially in the face of difficulty, has been confirmed in several empirical studies. For example, Helmke (1987) found that students' math-ability perceptions at the end of fifth grade had a positive impact on the quality of students' later efforts (e.g., on their perseverance and on their active engagement during instruction in sixth grade). Brown and Inouye's (1978) data indicated that the higher students' expectancies were concerning their ability to solve anagrams, the longer they persisted on anagrams for which they were unable to find solutions. Likewise, Hallerman and Meyer (1978, cited in Meyer, 1987) found that perceived ability was strongly

---

This research was made possible by a National Institute of Mental Health (NIMH) traineeship (MH187372) and by support from the Center for Research on Effective Schooling for Disadvantaged Students, a national research and development center funded by the Office of Educational Research and Improvement (OERI), U.S. Department of Education. The opinions presented are those of Douglas J. Mac Iver, Deborah J. Stipek, and Denise H. Daniels, and no official agreement by NIMH or OERI should be inferred.

We are indebted to the teachers and students in the Los Angeles Unified School District who participated in this research project.

Correspondence concerning this article should be addressed to Douglas Mac Iver at the Center for Research on Effective Schooling for Disadvantaged Students, Johns Hopkins University, 3505 North Charles Street, Baltimore, Maryland 21218.

Copyright 1991 by the American Psychological Association. Reprinted by permission.

predictive of the persistence of teenage students on insoluble achievement tasks, regardless of whether the tasks were portrayed as normatively easy or normatively difficult. Students who perceived their ability for the achievement task as high exhibited high persistence at both "easy" and "difficult" tasks. Furthermore, there is evidence that attributing one's learning difficulties to insufficient ability leads to decreased persistence (Andrews & Debus, 1978; Diener & Dweck, 1978; Licht, Kistner, Ozkaragoz, Shapiro, & Clausen, 1985; Weiner, 1979). For instance, Licht et al. (1985) found that, for both learning-disabled and non-learning-disabled children, the tendency to attribute one's failures to insufficient ability was negatively related to persistence on a reading task.

Although ability perceptions, especially as they determine one's expectancy for success on an achievement task, have been central in theories of achievement motivation, the value of a task to the individual is also assumed to influence his or her effort on that task. In Atkinson's (1957) theory, *task value* is narrowly defined in terms of the incentive value of success—the amount of pride one expects to experience if one succeeds. Difficult tasks are assumed to have higher incentive value than easy tasks. More recently, Parsons and Goff (1980), Reuman (1986), and others (e.g., Feather, 1988) have suggested that there are other reasons for valuing an achievement activity in addition to the pride one feels if one succeeds. According to Reuman (1986), these include "the inherent, immediate enjoyment one gets from developing, mastering or using a skill involved in the activity" (p. 92), that is, the *intrinsic or interest value* of the activity, and "the importance of the activity for some future goal" (p. 93), that is, the *utility value* of the activity.

Different relationships between expectancies and task values might be predicted depending on how one defines these two constructs. In the classic experimental research, expectancies for success were defined in terms of the inverse of the normative difficulty of the task, and only the incentive value of the task was considered. Expectancies for success (P_s) and the incentive value of success (I_s) were assumed to be perfectly inversely related: that is, $P_s = 1 - I_s$. In contrast, we propose that when expectancies for success are defined in terms of ability perceptions and the value of a task is defined in terms of its intrinsic and utility value, the relationship between expectancies and values will be strongly positive.

Our proposal is similar to Ryan, Connell, and Deci's (1985) proposition that "any event that enhances perceived competence will tend to enhance intrinsic motivation, while those that facilitate the perception of incompetence will diminish intrinsic motivation" (p. 17). In support of this proposition, Ryan et al. (1985) cited studies indicating that students who are provided with positive performance feedback concerning their competence on a task display higher levels of intrinsic motivation for the task than do students who don't receive performance feedback (Harackiewicz, 1979; Ryan, Chandler, Connell, & Deci, 1983). Furthermore, Vallerand and Reid (1982) reported evidence from a path analysis that suggests a causal link between feelings of competence in an activity and intrinsic motivation for that activity. In another study involving causal modeling techniques, Harter and Connell (1984) found that the structural equation models that best fit their data specify that pupils who evaluate their academic competence positively are more likely than others to be intrinsically motivated to engage in academic tasks.

There are at least two reasons to expect that ability perceptions also have a moderate positive effect on students' perceptions of the utility value of a course. First, students who believe that they are unable to master the knowledge and skills taught in a course may understandably question the course's usefulness to them (e.g., "If I can't master it, how will it help me in the future?"). Second, students tend to select career goals that require those talents that they think they have rather than those talents that they think they don't have. As a result, those activities that are perceived by students as useful in helping them reach their long-range goals also tend to be those activities at which they feel at least moderately talented.

In addition to the determinants of effort already mentioned, theories of achievement motivation emphasize that extrinsic pressures for achievement also influence effort (Atkinson, 1964; Ryan et al., 1985). For example, one might study hard in an attempt to please one's parents, even if one's ability perceptions or value perceptions in a course are low. Therefore, one issue we examined in this study is whether increases in the perceived importance of extrinsic pressures lead to increases in student effort.

In summary, effort on school tasks is assumed to be affected by ability perceptions, task-value perceptions, and perceptions of extrinsic pressures. In addition, higher ability perceptions are expected to lead to higher task-value perceptions. In our study, we tested these hypotheses. We also examined the relative importance of these factors in influencing effort and the nature (e.g., direct vs. indirect) of their effects.

In the analyses that follow, we evaluate four alternative models concerning the determinants of student effort in a course during junior high school and senior high school. Each model incorporates specific hypotheses concerning the causal relations among the following five correlated factors: (a) Change in Effort, (b) Change in Self-Concept of Ability, (c) Change in the Intrinsic Value of the Subject Matter, (d) Change in the Utility Value of the Course, and (e) Change in the Importance of Extrinsic Pressure for Achievement. Each model assumes that change in the importance of extrinsic pressures is a direct cause of change in effort. The models differ with regard to the roles in influencing effort that are attributed to changes in self-concept of ability, intrinsic value, and utility value. For example, one model assumes that the causal linkage between self-concept of ability and effort is a direct one. We compare this model with a model that assumes that the causal link between self-concept of ability and effort is mediated entirely through intrinsic value and utility value. A third model posits both direct and indirect causal links. Finally, we compare each of these models with a model that denies self-concept of ability a causal role in influencing effort; in this alternative model, change in self-concept of ability is viewed as a consequence rather than as a cause of changes in intrinsic value, utility value, and effort.

We included (and analyzed separately) data from both junior high school and senior high school students because the role of particular factors may differ in the middle grades and the high school grades. Utility value was expected to be more strongly associated with effort for high school students (who will confront occupational choices and endeavors sooner than will junior high school students), and the importance of extrinsic pressures was expected to be more strongly associated with effort for junior high school students (who are more influenced by a desire to please parents than are high school students).

Method

Students from two junior high schools and two senior high schools in southern California were recruited for this study. Each school had a diverse student body that included students from a broad spectrum of social classes and ethnic groups. Initially 23 teachers volunteered for the study, but 3 changed their minds and dropped out before the study was completed. Of the teachers who completed the study, 40% were math teachers, 20% were English teachers, 15% were science teachers, and the remaining 25% taught social studies or elective subjects (e.g., Spanish, computer education, photography).

Participating teachers distributed parental permission forms to the students in one or two of their classes. Students who returned signed permission forms were allowed to participate in the study. The number of participating students varied considerably from classroom to classroom because some teachers issued daily reminders to bring back the permission slips before "questionnaire day," whereas others issued few or no reminders. The overall student participation rate was about 60%. Of the participating students, 46% were White, 23% were Hispanic, 13% were Black, and 18% were from other ethnic groups; 46% of the students who participated were boys and 54% were girls.

To obtain multiple indicators of change, several measures of each construct were included on a survey questionnaire that was administered to students twice: once within the first 2 weeks of the semester and once at the end of the semester. A total of 322 students (167 junior high school students and 155 senior high school students) filled out both the beginning-of-semester and the end-of-semester questionnaires. Teachers were asked to complete an assessment of each participating student's effort both at the beginning and at the end of the semester. These effort ratings by teachers were obtained for 282 of the 322 students.

The Appendix lists the items that were used to measure the five constructs assessed. With the exception of two effort items, all items have a response scale ranging from 1 to 7 with various anchors, as indicated in the Appendix. Each item focuses on the subject area of the specific course in which students were given the questionnaire. To maximize the construct validity of the change in effort factor, we combined information from two independent sources (student self-report items and teachers' ratings) in measuring effort changes (see Nunnally, 1978, p. 98). For each item in each factor, the difference between a student's or teacher's rating at the end of the semester and at the beginning of the semester was used as an estimate of change. These change estimates were the basic data used in the analyses in this article.[1]

Results

Overview of Analysis Strategy

We used confirmatory factor analysis to assess the adequacy of the proposed five-factor measurement model. Then we conducted LISREL analyses to test the adequacy of several alternative covariance structure models (models specifying not only the factor structure but also the causal relations among factors). In both types of analyses, we evaluated the adequacy of hypothesized models by examining the congruence between the covariance matrix generated by the hypothesized model and the observed covariance matrix. We used the Tucker-Lewis index (TLI) to assess whether the overall fit (between the covariances generated by the hypothesized model and the observed covariances) was good enough to support the model (Tucker & Lewis, 1973). The TLI is the only widely used goodness-of-fit index that is relatively independent of sample size (Marsh, Balla, & McDonald, 1988). Although there is not universal agreement on what constitutes "good" fit, a value of .90 or better on the TLI is usually considered acceptable (Marsh et al., 1988, p. 393). A TLI of .90 indicates that the proposed model improves the null model by 90% of the amount one would expect from a model that is precisely true.

Confirmatory Factor Analysis

In the confirmatory factor analysis, each indicator of change listed in the Appendix was constrained to load only on the factor that it was designed to measure. The results of the analysis indicated that the hypothesized factor structure fits the data well (χ^2/df ratio = 1.30, TLI = .93). Furthermore, the factor loadings (given in the Appendix) and factor variances were large and statistically significant. (We adopted a .05 probability level for all significance tests reported in this article.) As anticipated, all correlations among factors (see Table 1) were positive and, except for the correlation between change in the importance of extrinsic pressures and change in self-concept of ability, significant.

These means and standard deviations of the five factors were estimated with Bollen's (1989, pp. 306–311) method (See Table 2). These estimates reveal that, on average, there are negative within-semester changes in effort, in the perceived importance of extrinsic pressures, in self-concept of ability, and in intrinsic value. On the other hand, the average within-semester change in students' perceptions of the utility value of their coursework is slightly positive. Finally, the standard deviations in Table 2 indicate that there is considerable variation among students in the within-semester changes that they exhibit.

Covariance Structure Models

The confirmatory factor analysis supported the hypothesis that the within-semester change scores computed from stu-

---

[1] Even though observed difference scores provide unbiased estimates of true change, many authors have criticized the use of these scores (e.g., Bereiter, 1963; Bohrnstedt, 1969; Kessler, 1977; Linn & Slinde, 1977; O'Connor, 1972). However, as Willett (1988, p. 367) concluded, recent methodological research has revealed that the purported deficiencies of difference scores "are *perceived* rather than *actual*, *imaginary* rather than *real* (Rogosa, Brandt, & Zimowski, 1982; Rogosa & Willett, 1983, 1985; see also Zimmerman, Brotohusodo, & Williams, 1981; Zimmerman & Williams, 1982)."

Table 1
Factor Correlations

Factor	1	2	3	4	5
1. ΔINT	—				
2. ΔEXT	.24	—			
3. ΔUTI	.32	.22	—		
4. ΔSCA	.73	.15	.34	—	
5. ΔEFF	.52	.36	.31	.58	—

Note. ΔINT = Change in the Intrinsic Value of the Subject Matter; ΔEXT = Change in the Importance of Extrinsic Pressures for Achievement; ΔUTI = Change in the Utility Value of the Course; ΔSCA = Change in Self-Concept of Ability; ΔEFF = Change in Effort.

dents' responses to the items in the Appendix measure the five correlated factors that they were intended to measure. Therefore, in the covariance structure analyses, we used the hypothesized five-factor structure as the measurement model. In these analyses, each factor was measured by its four best indicators. The metric of each factor was set (with the reference indicators listed in Table 2) to measure within-semester change on a 13-point scale; the maximum possible positive change was +6 and the maximum possible negative change was −6.[2]

As described earlier, four alternative causal models were tested. In the following sections, we describe the results of analyses conducted to test the fit of each model to the empirical data.

Model 1. The causal relations specified in Model 1 (depicted in Figure 1) reflect our hypothesis that when students lower or raise their estimate of ability in a subject, their effort in the subject is affected, as is their valuing of the subject (e.g., "If I discover I'm good at a subject, I'm more willing to put forth effort in the subject, and I'm more likely to perceive the subject to be interesting and useful"). In addition, Model 1 specifies that changes in the importance of extrinsic pressures (e.g., an increased desire to please parents or to obtain a good grade) affect effort.

Unstandardized maximum-likelihood parameter estimates for Model 1 were obtained separately for junior high school students and senior high school students with simultaneous multisample analysis in LISREL VI. These estimates are reported in Figure 1.[3] For each path, the estimate for junior high school students is listed first (to the left of the slash). Model 1 explains 70% of the variance in change in effort levels in junior high school students and 34% of this variance in senior high school students. In both junior high school and senior high school, a within-semester change in students' self-concept of ability has a substantial and statistically significant impact on their effort in that course. For example, for junior high school students, an increase in self-concept of ability of 1 point is associated with an increase in effort of 0.7 points. Change in the importance of extrinsic pressures has a significant impact on change in effort in junior high school but not in senior high school. In both junior high school and senior high school, change in self-concept of ability is positively associated with change in intrinsic value and with change in utility value. Change in self-concept of ability explains over 50% of the variance in change in intrinsic value and over 10% of the variance in change in utility value in both samples. Finally, Model 1 fits the data well (χ^2/df ratio = 1.15, TLI = .94).

Model 2. Model 2 (depicted in Figure 2) differs from Model 1 in that it specifies that some of the effect of change in self-concept of ability on change in effort is indirect. That is, in addition to directly affecting effort, change in ability perceptions indirectly affects effort by causing changes in the perceived intrinsic value and utility value of the subject matter.

As in Model 1, the parameter estimates in Model 2 are consistent with the following assertions: (a) within-semester-changes in students' course-related ability perceptions affect both students' effort and their valuing of the subject matter; and (b) in junior high school, change in the perceived importance of parents and grades as extrinsic motivators leads to changes in effort. In addition, there are several indications from the output associated with Model 2 that the direct effect of change in ability perceptions on change in effort may be more important than its indirect effects. First, Model 2 (which contains the indirect effects) does not fit the data significantly better than Model 1, $\Delta\chi^2 = 3.73$, $\Delta df = 4$, *ns*. Second, the parameter estimates from Model 2 suggest that the direct effect of change in self-concept of ability on change in effort is larger than the sum of the indirect effects: the estimated direct effect is 0.56 in junior high school and 0.41 in senior high school, whereas the sum of the indirect effects is only 0.14 in junior high school and 0.16 in senior high school.

It should be noted, however, that the high correlation between change in self-concept of ability and change in intrinsic value (.73) makes it difficult to estimate precisely the relative magnitudes of the direct effect and the indirect effects. Because of this high correlation, the parameter estimates for the direct effect and one of the indirect effects are highly correlated (−.80). One result of the correlation between these two parameter estimates is that the standard error associated with the effect of change in self-concept of ability on change in effort is 1.7 times higher in Model 2 than it is in Model 1. For this reason, in Model 2 the significance level associated with this effect does not reach conventional levels of significance.

Model 3. Another way of testing the importance of the direct effect of change in self-concept of ability on change in effort is to compare the fit of a model that contains the direct

---

[2] When interindividual differences in true change are small, the difference between two observed measures tends to be less reliable than either individual measure. In regression analyses and other traditional statistical techniques, unreliable measures increase bias in parameter estimates involving those measures as predictors and inflate standard errors of estimate. However, if change in each construct is measured by several difference scores, covariance structure modeling techniques can be used (as we use them here) to obtain parameter estimates that are unbiased by the random measurement error in the difference scores.

[3] There are three types of parameter estimates reported in Figures 1–4, each represented by a different type of line. A curved line with arrowheads at both ends represents the covariance between exogenous factors. The direct effect of one factor on another is represented by a straight line (or by connected line segments) with one arrowhead to show the assumed direction of causation. Finally, an arrowhead without a tail represents specification error (error in equations or omitted variables).

Table 2
Reference Indicators for Five Change-Score Factors and Estimated Means and Standard Deviations for These Factors

Factor	Wording of reference indicator	Maximum possible change	*M* change	*SD*
ΔINT	How excited are you to learn about this subject matter?	+6 (from *not at all excited* to *very excited*)	−0.17	1.12
ΔEXT	Is doing as well as your parents expect you to do in this class important to you?	+6 (from *not at all important to me* to *very important to me*)	−0.43	1.12
ΔUTI	I am taking this class because it helps prepare me for a job.	+6 (from *not an important reason at all* to *a very important reason*)	+0.07	1.47
ΔSCA	How good are you in this subject?	+6 (from *not good at all* to *very good*)	−0.22	0.91
ΔEFF	How hard are you working to learn about this subject?	+6 (from *not hard at all* to *as hard as I can*)	−0.55	0.91

Note. ΔINT = Change in the Intrinsic Value of the Subject Matter; ΔEXT = Change in the Importance of Extrinsic Pressures for Achievement; ΔUTI = Change in the Utility Value of the Course; ΔSCA = Change in Self-Concept of Ability; ΔEFF = Change in Effort.

effect (e.g., Model 2) with a nested model that eliminates the direct effect (Model 3 in Figure 3). The fit of Model 3 is significantly worse than the fit of Model 2: $\Delta\chi^2 = 7.18$, $\Delta df = 2$, $p < .05$. Thus, the evidence suggests the existence of the direct effect.

In conclusion, the results from the first three models suggests that (a) change in ability perceptions has an important direct effect on change in effort; (b) change in the perceived importance of extrinsic pressures for achievement has a significant positive impact on effort in junior high school but not in senior high school; and (c) as students' ability perceptions in a subject increase, not only do they try harder but they also enjoy the subject more and perceive the subject as more useful.

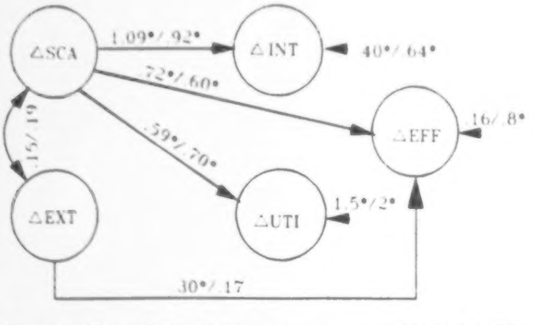

Figure 1. Unstandardized parameter estimates for Model 1. (ΔSCA = Change in Self-Concept of Ability; ΔINT = Change in the Intrinsic Value of the Subject Matter; ΔEFF = Change in Effort; ΔEXT = Change in the Importance of Extrinsic Pressures for Achievement; ΔUTI = Change in the Utility Value of the Course. Values to the left of the slash are for junior high school students, and values to the right of the slash are for senior high school students. There are three types of parameter estimates, each represented by a different type of line. A curved line with arrowheads at both ends represents the covariance between exogenous factors. The direct effect of one factor on another is represented by a straight line [or by connected line segments] with one arrowhead to show the assumed direction of causation. Finally, an arrowhead without a tail represents specification error [error in equations or omitted variables]. An asterisk indicates a coefficient that is greater than or equal to 1.96 times its standard error. Tucker-Lewis index = .94.)

Models 1, 2, and 3 assume that the effect of change in self-concept of ability on change in effort does not depend on the level of students' ability perceptions at the beginning of the course. This assumption was confirmed with multiple regression analyses. In these analyses, students were first categorized into quartiles on the basis of their ability perceptions at the beginning of the semester: low perceived ability (an average rating of 4 or less), moderate perceived ability (an average rating greater than 4 but less than or equal to 5), high perceived ability (an average rating greater than 5 but less than or equal to 5.75), and very high perceived ability (an average rating greater than 5.75). Then we estimated the simple regression of change in effort on change in self-concept of ability separately for each quartile and tested whether the regression coefficients in the separate regression equations were significantly different from each other (i.e., whether or not the regression lines were parallel). These tests revealed that the positive impact of increased ability perceptions on effort is not significantly different for students who begin the year with perceptions of very high, high, moderate, or low ability: in the junior high school sample, $F(3, 142) = 1.60$, $p = .19$, $MS_e = .69$, and in the senior high school sample, $F(3, 136) = 1.46$, $p = .23$, $MS_e = .73$.

Model 4: A Supplementary Analysis to Test an Alternative Explanation. The first three models all assume that changes in self-concept of ability lead to changes in effort. An alternative view is that ability perception change is a consequence rather than a cause of effort change. Perhaps, for example, (a) a student discovers that a subject is more interesting or more enjoyable than he or she originally thought, (b) the student's increased enjoyment prompts him or her to work harder, and (c) the student's hard work increases his or her competence and consequently his or her ability perceptions. This alternative causal sequence is specified in Model 4 (Figure 4). Model 4 fits the data less well than do Models 1, 2, and 3: for Model 4, χ^2/df ratio = 1.24, TLI = .91.[4]

---

[4] At the suggestion of an anonymous reviewer, we also tested the explanatory power and overall fit of several *state-dependence* models that specify that students' beginning-of-semester perceptions (of intrinsic value, utility value, ability, and extrinsic pressures) determine subsequent effort changes. These models had substantially poorer explanatory power and poorer overall fit than the models reported in this article.

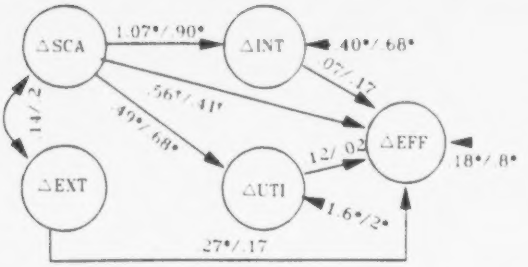

Figure 2. Unstandardized parameter estimates for Model 2. (∆SCA = Change in Self-Concept of Ability; ∆INT = Change in the Intrinsic Value of the Subject Matter; ∆EFF = Change in Effort; ∆EXT = Change in the Importance of Extrinsic Pressures for Achievement; ∆UTI = Change in the Utility Value of the Course. Values to the left of the slash are for junior high school students, and values to the right of the slash are for senior high school students. There are three types of parameter estimates, each represented by a different type of line. A curved line with arrowheads at both ends represents the covariance between exogenous factors. The direct effort of one factor on another is represented by a straight line [or by connected line segments] with one arrowhead to show the assumed direction of causation. Finally, an arrowhead without a tail represents specification error [error in equations or omitted variables]. An asterisk indicates a coefficient that is greater than or equal to 1.96 times its standard error. A dagger indicates a coefficient that is greater than or equal to 1.65 times its standard error. Tucker-Lewis index = .94.)

Mean Differences in Within-Semester Change Among Courses of Different Types

As reported earlier, averaging across the different types of courses in our sample, we found that students showed negative within-semester changes in effort, in the perceived importance of extrinsic pressure, in self-concept of ability, and in the intrinsic value of the subject matter. Although our study was not designed to allow definitive statements about differences among courses of different types, the data in Table 3 indicate that—at least in our sample—negative within-semester changes tend to be more prevalent and more pronounced in math courses than in other types of courses. Mean change in intrinsic value, change in self-concept of ability, and change in effort scores of students in math courses were significantly lower than those of students in every other course type. Whereas students in math courses did not differ from students in science, social studies, or elective courses in their increasingly negative perceptions of the course's utility value, students in English courses came to view English as significantly more useful as the semester progressed. There were no significant course-type differences on change in the importance of extrinsic pressures, $F(3, 303) = 1.33$, $p = .26$, $MS_e = 0.89$; students in all courses showed negative changes on this factor.

Although there were significant mean differences among course types in change in intrinsic value, change in utility value, change in self-concept of ability, and change in effort, this does not imply that effort changes are related differently to task value or to ability perception changes in courses of different types. On the contrary, regression analyses reveal that course type does not affect the relations of change in effort with change in intrinsic value, change in utility value, change in self-concept of ability, or change in the importance of extrinsic pressures. For example, we estimated the simple regression of effort change on ability-perception change separately for math, English, science, and other courses. Then we tested for significant differences in the estimated regression coefficients. These tests revealed that the positive impact of increased ability perceptions on effort is not significantly different for students in different courses, $F(3, 246) = 0.87$; $p = .46$, $MS_e = 0.55$.

Discussion

As mentioned earlier, in virtually every cognitive theory of motivation, ability perceptions are assumed to affect student effort and thus to have practical, educational importance. In an implicit endorsement of these theories, many of the task forces and commissions attempting to reform the schools attended by young adolescents have emphasized the importance of providing young adolescents with experiences and interactions that help them develop a self-image of intellectual competence (e.g., Carnegie Task Force on the Education of Young Adolescents, 1989; Children's Defense Fund, 1988; Maryland Task Force on the Middle Learning Years, 1989). The results of this study suggest that this emphasis on perceptions of ability may be worthwhile. Models 1 and 2 (which explicitly assume that within-semester change in ability perceptions is a direct determinant of within-semester change in

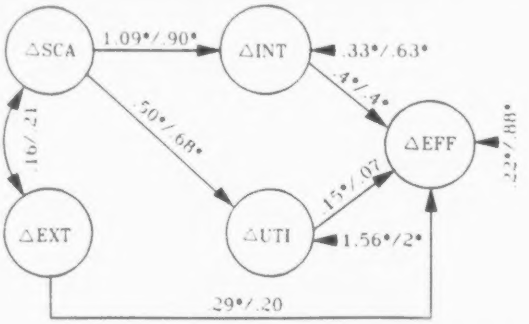

Figure 3. Unstandardized parameter estimates for Model 3. (∆SCA = Change in Self-Concept of Ability; ∆INT = Change in the Intrinsic Value of the Subject Matter; ∆EFF = Change in Effort; ∆EXT = Change in the Importance of Extrinsic Pressures for Achievement; ∆UTI = Change in the Utility Value of the Course. Values to the left of the slash are for junior high school students, and values to the right of the slash are for senior high school students. There are three types of parameter estimates, each represented by a different type of line. A curved line with arrowheads at both ends represents the covariance between exogenous factors. The direct effect of one factor on another is represented by a straight line [or by connected line segments] with one arrowhead to show the assumed direction of causation. Finally, an arrowhead without a tail represents specification error [error in equations or omitted variables]. An asterisk indicates a coefficient that is greater than or equal to 1.96 times its standard error. Tucker-Lewis index = .93.)

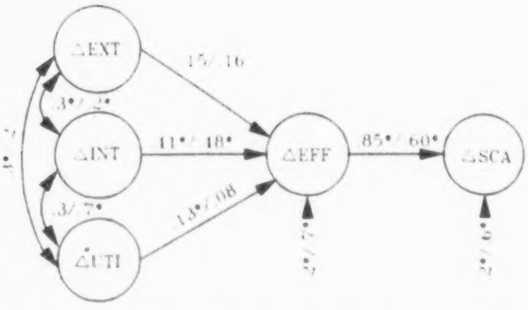

Figure 4 Unstandardized parameter estimates for Model 4. (ΔEXT = Change in the Importance of Extrinsic Pressure for Achievement; ΔINT = Change in the Intrinsic Value of the Subject Matter; ΔEFF = Change in Effort; ΔSCA = Change in Self-Concept of Ability; ΔUTI = Change in the Utility Value of the Course. Values to the left of the slash are for junior high school students, and values to the right of the slash are for senior high school students. There are three types of parameter estimates, each represented by a different type of line. A curved line with arrowheads at both ends represents the covariance between exogenous factors. The direct effect of one factor on another is represented by a straight line [or by connected line segments] with one arrowhead to show the assumed direction of causation. Finally, an arrowhead without a tail represents specification error [error in equations or omitted variables]. An asterisk indicates a coefficient that is greater than or equal to 1.96 times its standard error. Tucker-Lewis index = .91.)

effort) fit the data better than do the alternative models considered here. In other words, the results are consistent with the claim that, by reducing the number of students who believe that they are "not good" in a subject, teachers can increase the number of students who work near their potential.

The findings suggest that increasing students' perceptions of ability will achieve another important goal: for students to value the subject they are learning. Students whose ability perceptions in a subject increase find the subject to be more interesting and more useful; conversely, students devalue subjects that they do not believe they have the ability and skills to master. The effect of ability perceptions on values may have long-term implications. For example, greater valuing of a subject may result in students seeking further learning opportunities in that subject area.

The results of this study support Eccles & Wigfield's (1985a) contention that expectancies (as measured by self-concept of ability) and task values are positively related in naturally occurring achievement settings. Students come to value those subjects at which they believe they can succeed. This relationship is the opposite of the relationship between expectancies and incentive value proposed by Atkinson (1964). Clearly, the distinction between the incentive value of a task (defined narrowly in terms of anticipated pride accompanying success) and task values (more broadly construed and applied to natural achievement contexts) has important theoretical and practical implications.

Previous research has found that students' beliefs concerning the value of academic subjects influence their course enrollment decisions (e.g., Chipman, Brush, & Wilson, 1985; Eccles, Adler, & Meece, 1984). The results of this study suggest, however, that once a student is enrolled in a course, changes in value perceptions may not affect effort. Model 1 (which assumes that effort change and changes in the valuing of a subject are correlated only because of their common dependence on changes in ability perceptions) fits the data quite well. Furthermore, in Model 2, the estimated effects of intrinsic value and utility value changes on effort were small and insignificant. Only Models 3 and 4—which omit the direct effect of ability perception change on effort change and which fit worse than Models 1 and 2—yield any significant effects of value-of-subject changes on effort changes. A useful goal of future research would be to delineate the circumstances under which students' perceptions of the subjective value of a task play an important role in determining effort, after controlling for ability perceptions.

For many junior high school students, the desire to please one's parents by getting a good grade is an important reason for putting forth effort in a class. Increases in the perceived importance of this type of extrinsic pressure for achievement were significantly associated with increased effort among junior high school students but not among senior high school students. This developmental difference probably reflects the declining importance of parental norms and pressures in influencing students' achievement behavior from early to late adolescence (e.g., Montemayor, 1986). It is possible that by high school peer-related extrinsic pressures supersede parental pressures. Of course, peer-based pressures for or against achievement may be important even during the upper-elementary and middle grades (e.g., Slavin, 1986). The relative

Table 3

Means and Standard Deviations for Five Change-Score Composites by Class Type

	Math		English		Science		Other	
Composite	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
ΔINT	−0.40	1.09	+0.03	1.03	+0.20	1.14	−0.08	0.96
ΔEXT	−0.45	0.88	−0.16	0.78	−0.33	1.29	−0.29	0.97
ΔUTI	−0.26	1.29	+0.44	1.09	−0.15	1.44	−0.03	1.11
ΔSCA	−0.42	1.16	−0.08	0.82	+0.03	0.76	−0.06	0.85
ΔEFF	−0.63	0.97	−0.12	0.67	−0.29	0.77	−0.37	0.72

Note. ΔINT = Change in the Intrinsic Value of the Subject Matter; ΔEXT = Change in the Importance of Extrinsic Pressures for Achievement; ΔUTI = Change in the Utility Value of the Course; ΔSCA = Change in Self-Concept of Ability; ΔEFF = Change in Effort.

effects of perceived peer and parent pressure on effort would be useful to examine in future studies.

It is noteworthy that the best model tested in this study explained twice as much of the variance in effort for the junior high school students (70%) as for the senior high school students (34%). Apparently the four psychological factors assessed in the study give a more complete picture of the determinants of effort on school tasks for the younger adolescents. Further research needs to examine additional factors that may influence older students' effort on school tasks. We suspect that competing activities, such as athletic training, jobs, and peer relationships, affect time spent on schoolwork for senior high school students more than for junior high school students.

Educators are searching continually for "promising practices" to improve students' motivation and performance. By adding to our knowledge of the relative importance of factors that affect effort and of the relations among these factors, the results of this study may help teachers select strategies to increase student effort. The results suggest that the most fruitful approach to increasing student effort may involve altering curriculum and instruction, task structures, grouping policies, and evaluation practices to reduce the proportion of students who perceive themselves as having little academic ability (Mac Iver, 1988; Mac Iver & Epstein, 1990). However, raising students' confidence in their abilities is a complicated and difficult task. Seemingly positive teacher behaviors that are motivated by a desire to protect the self-concept of low achievers often inadvertently play a role in damaging this self-concept. For example, Graham & Barker (1990, p. 7) caution that "praise for success at easy tasks, the absence of blame for failure at such tasks, and affective displays of sympathy or compassion can communicate to the recipients of this feedback that they are low in ability (Barker & Graham, 1987; Graham, 1984; Meyer et al., 1979; see also Weiner, Graham, Taylor, & Meyer, 1983)." Increased perceptions of competence cannot be effectively achieved by setting unchallenging standards for success. On this point, we agree with Atkinson that success easily achieved engenders little pride (and also fails to increase self-confidence).

One promising approach to raising students' ability perceptions is to alter classroom evaluation and recognition practices so that success is defined in terms of individually referenced (and personally challenging) improvement goals. Mac Iver (1990) described a program for middle-grade classrooms that follows this approach. The program helps students to set individualized, short-range improvement goals that are challenging but doable. As young adolescents observe their progress in obtaining these goals, many who have reached the premature conclusion that they will "never be good at schoolwork" may develop a renewed confidence in their academic ability and a renewed enthusiasm for learning. A multiyear evaluation study of this program is currently under way.

In addition to building students' self-confidence in their ability, should teachers also stress the utility value of mastering course content and strive to make school tasks intrinsically interesting? These are undoubtedly good teaching strategies. Attempts to make the content of a course more clearly useful to students and to make assignments in a class more interesting may increase students' enjoyment of a course and positive attitudes toward school in general, and they may have long-term effects on the degree to which students seek further learning opportunities in that subject area. However, this study suggests that strategies directed toward ensuring that all students develop faith in their ability may have a greater effect on student's effort and attitudes at least in the short run (e.g., during the course of a semester), than would strategies directed toward increasing the utility value and intrinsic interest of a course. As Eccles and Wigfield (1985b) have argued,

> One of the most important motivational questions facing a student is "Can I succeed at this task if I choose to try?" . . . . If the answer is yes, then a student will, at least, move on to the next question—"Do I want to?" If the answer is no, then the student will, in all likelihood, give up. (p. 188)

Classroom practices that increase the number of children who gain confidence in their ability may help create a success-prone cycle in individual children. That is, increases in effort (resulting from increases in perceived competence) may lead students to succeed more frequently. This increased success may prompt further increases in confidence and effort, thus creating a success-prone cycle. On the other hand, despite the importance of ability perceptions in motivating effort, heightened ability perceptions will be of little use "unless accompanied by the strategic knowledge that is essential to direct the energy to appropriate ends" (Nickerson, 1988, p. 26; see also Borkowski, Carr, Rellinger, & Pressley, in press). Thus, one critical component of effective confidence-building programs may be the provision of direct instruction in metacognitive strategies.

One limitation of this study is its inability to identify and analyze differences among different socioeconomic and ethnic groups in the determinants of effort changes. We did not collect information on the socioeconomic status of individual students, and we have insufficient numbers of students within each minority group to permit a LISREL analysis of ethnic group differences. Simple regressions conducted within each ethnic group separately indicated that ability perception change is strongly related to effort change within every ethnic group. Nevertheless, future studies in which socioeconomic status measures are collected and in which larger minority group samples are evaluated may find that our best model does not satisfactorily fit the data from certain groups.

References

Abramson, L. Y., Seligman, M. E. P., & Teasdale, J. D. (1978). Learned helplessness in humans: Critique and reformulation. *Journal of Abnormal Psychology, 87,* 49–74.

Andrews, G. R., & Debus, R. L. (1978). Persistence and the causal perception of failure: Modifying cognitive attributions. *Journal of Educational Psychology, 70,* 154–166.

Atkinson, J. (1957). Motivational determinants of risk-taking behavior. *Psychological Review, 64,* 359–372.

Atkinson, J. (1964). *An introduction to motivation.* Princeton, NJ: Van Nostrand.

Atkinson, J. W., & Birch, D. (1978). *An introduction to motivation* (2nd ed.). New York: Van Nostrand.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191–215.

Bandura, A., & Cervone, D. (1983). Self-evaluative and self-efficacy mechanisms governing the motivational effects of goal systems. *Journal of Personality and Social Psychology, 45*, 1017–1028.

Bereiter, C. (1963). Some persisting dilemmas in the measurement of change. In C. W. Harris (Ed.), *Problems in measuring change*. Madison: University of Wisconsin Press.

Bohrnstedt, G. W. (1969). Observations on the measurement of change. In E. F. Borgatta (Ed.), *Sociological Methodology, 1969*. San Francisco: Jossey-Bass.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Borkowski, J. G., Carr, M., Rellinger, E., & Pressley, M. (1990). Self-regulated cognition: Interdependence of metacognition, attributions, and self-esteem. In B. F. Jones & L. Idol (Eds.), *Dimensions of thinking and cognitive instruction* (Vol. 1, pp. 53–92). Hillsdale, NJ: Erlbaum.

Brown, I., & Inouye, K. (1978). Learned helplessness through modeling: The role of perceived similarity in competence. *Journal of Personality and Social Psychology, 36*, 900–908.

Carnegie Task Force on the Education of Young Adolescents. (1989). *Turning points: Preparing American youth for the 21st century*. New York: Carnegie Council on Adolescent Development of the Carnegie Corporation.

Children's Defense Fund. (1988). *Making the middle grades work*. Washington, DC: Author.

Chipman, S. F., Brush, L. R., & Wilson, D. M. (Eds.). (1985). *Women and mathematics: Balancing the equation*. Hillsdale, NJ: Erlbaum.

Covington, M., & Beery, R. (1976). *Self-worth and school learning*. New York: Holt, Rinehart & Winston.

Diener, C. I., & Dweck, C. S. (1978). An analysis of learned helplessness: Continuous changes in performance, strategy, and achievement cognitions following failure. *Journal of Personality and Social Psychology, 36*, 451–462.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41*, 1040–1048.

Eccles, J., Adler, T. F., & Meece, J. L. (1984). Sex differences in achievement: A test of alternative theories. *Journal of Personality and Social Psychology, 46*, 26–43.

Eccles, J., & Wigfield, A. (1985a). *Math achievement patterns: Social and psychological forces*. (Grant No. BNS-8510504). Washington, DC: National Science Foundation.

Eccles, J., & Wigfield, A. (1985b). Teacher expectations and student motivation. In J. B. Dusek (Ed.), *Teacher expectations* (pp. 185–226). Hillsdale: NJ: Erlbaum.

Feather, N. T. (1961). The relationship of persistence at a task to expectations of success and achievement-related motives. *Journal of Abnormal and Social Psychology, 63*, 552–561.

Feather, N. T. (1988). Values, valences, and course enrollment: Testing the role of personal values within an expectancy–valence framework. *Journal of Educational Psychology, 80*, 381–391.

Graham, S., & Barker, G. P. (1990). The down side of help: An attributional-developmental analysis of helping behavior as a low-ability cue. *Journal of Educational Psychology, 82*, 7–14.

Harackiewicz, J. (1979). The effects of reward contingency and performance feedback on intrinsic motivation. *Journal of Personality and Social Psychology, 37*, 1352–1363.

Harter, S., & Connell, J. (1984). A comparison of alternative models of the relationships between academic achievement and children's perceptions of competence, control, and motivational orientation. In J. Nicholls (Ed.), *The development of achievement-related cognitions and behavior* (pp. 219–250). Greenwich, CT: JAI Press.

Helmke, A. (1987). *Mediating processes between children's self-concept of ability and mathematics achievement: A longitudinal study* (Paper No. 6). Munich, Bavaria, Germany: Max Planck Institute for Psychological Research.

Kessler, R. C. (1977). The use of change scores as criteria in longitudinal survey research. *Quality and Quantity, 11*, 43–66.

Kukla, A. (1978). An attributional theory of choice. *Advances in Experimental Social Psychology, 11*, 113–144.

Licht, B. G., Kistner, J. A., Ozkaragoz, T., Shapiro, S., & Clausen, L. (1985). Causal attributions of learning disabled children: Individual differences and their implications for persistence. *Journal of Educational Psychology, 77*, 208–216.

Linn, R. L., & Slinde, J. A. (1977). The determination of the significance of change between pre- and posttesting periods. *Review of Educational Research, 47*, 121–150.

Mac Iver, D. (1988). Classroom environments and the stratification of students' ability perceptions. *Journal of Educational Psychology, 80*, 495–505.

Mac Iver, D. (1990). *Altering evaluations, rewards, and recognitions to better motivate young adolescents: The Incentives for Improvement Program*. Baltimore: Johns Hopkins University Center for Research on Effective Schooling for Disadvantaged Students.

Mac Iver, D., & Epstein, J. L. (1990). *Responsive education in the middle grades: Teacher teams, advisory groups, remedial instruction, school transition programs, and report card entries* (Report No. 46). Baltimore: Johns Hopkins University Center for Research on Elementary and Middle Schools.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103*, 391–410.

Maryland Task Force on the Middle Learning Years. (1989). *What matters in the middle grades?* Baltimore: Maryland State Department of Education.

Meyer, W.-U. (1987). Perceived ability and achievement-related behavior. In F. Halisch & J. Kuhl (Eds.), *Motivation, intention, and volition*. Springer-Verlag: Berlin, Germany.

Montemayor, R. (1986). Developing autonomy: The transition of youth to adulthood. In G. K. Leigh & G. W. Peterson (Eds.) *Adolescents in families* (pp. 205–225). Cincinnati, OH: South-Western.

Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review, 91*, 328–346.

Nickerson, R. S. (1988). On improving thinking through instruction. *Review of Research in Education, 15*, 3–57.

Nunnally, J. C. (1978). *Psychometric Theory* (2nd ed.). New York: McGraw-Hill.

O'Connor, E. F. (1972). Extending classical test theory to the measurement of change. *Review of Educational Research, 42*, 73–98.

Parsons, J. E., & Goff, S. B. (1980). Achievement motivation: A dual modality. In L. J. Fyans (Ed.), *Recent trends in achievement motivation: Theory and research*. New York: Plenum Press.

Raynor, J. O., & Brown, E. T. (1985). Motivation at different stages of striving in a psychological career. In D. A. Klieber (Ed.), *Advances in motivation and achievement: Vol. 4. Motivation and adulthood* (pp. 121–167). Greenwich, CT: JAI Press.

Reuman, D. A. (1986). Motivational implications of ability grouping in sixth-grade mathematics: A strong inference approach to theories of achievement motivation (Doctoral Dissertation, University of Michigan, 1986). *Dissertation Abstracts International, 47*, 1315B.

Ryan, R. M., Chandler, C., Connell, J. P., & Deci, E. L. (1983, April). *Internalization and motivation: Some preliminary research and theoretical speculations*. Paper presented at the biennial meeting of the Society for Research in Child Development, Detroit, MI.

Ryan, R., Connell, J., & Deci, E. (1985). A motivational analysis of self-determination and self-regulation. In C. Ames & R. Ames (Eds.), *Research on motivation in education: Vol. 2. The classroom* *milieu* (pp. 13–51). San Diego, CA: Academic Press.

Slavin, R. E. (1986). *Using student team learning.* Baltimore: Johns Hopkins Team Learning Project.

Tucker, L. R., & Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*, 1–10.

Vallerand, R. J., & Reid, G. (1982). *On the causal effects of perceived competence on intrinsic motivation.* Unpublished manuscript, University of Guelph, Guelph, Ontario, Canada.

Weiner, B. (1979). A theory of motivation for some classroom experiences. *Journal of Educational Psychology, 71*, 3–25.

Weiner, B. (1986). *An attributional theory of motivation and emotion.* New York: Springer-Verlag.

Willett, J. B. (1988). Questions and answers in the measurement of change. *Review of Research in Education, 15*, 345–422.

Appendix

Questionnaire Items Administered at the Beginning and the End of a Semester to Measure Within-Semester Change in Five Factors

Item	Response scale anchors	Loading
	Change in the Intrinsic Value of the Subject Matter	
How excited are you to learn about this subject matter?	*Not at all excited; very excited*	1.06
How much do you enjoy learning about this subject?	*Not much at all; very much*	1.14
How much do you care about learning a lot about this subject?	*Don't care at all; care very much*	0.85
How much do you like working on the assignments in this class?	*Not at all; very much*	0.90
Do you do things for fun outside of class that are related to or have something to do with what you are learning about in this class?	*Never; yes, a lot*	0.57
	Change in the Importance of Extrinsic Pressures for Achievement	
When I work in this class, it is because I want to please my parents.	*Not at all a reason; a very important reason*	0.79
Is doing as well as your parents expect you to do in this class important to you?	*Not at all important to me; very important to me*	1.19
When I work in this class, it is because I want a good grade.	*Not at all a reason; a very important reason*	0.54
How important is it to your parent(s) that you get a good grade in this class?	*Not at all important; very important*	0.54
	Change in the Utility Value of the Course	
I am taking this class because it helps prepare me for a job.	*Not an important reason at all; a very important reason*	1.46
When I work in this class, it is because the knowledge and skills are useful in my life and/or for my future.	*Not at all a reason; a very important reason*	1.22
I am taking this class because I may need to know about this subject in the future.	*Not an important reason at all; a very important reason*	1.33
I am taking this class because it helps me do things I want to be able to do.	*Not an important reason at all; a very important reason*	0.96
How useful is what you learn in this class for a job you might want?	*Not at all useful; very useful*	1.02
I am taking this class because it helps me decide what career or job I want.	*Not an important reason at all; a very important reason*	1.14
How useful will what you learn in this class be for future classes you might take?	*Not at all useful; very useful*	0.72

Appendix (*continued*)

Item	Response scale anchors	Loading
	Change in Self-Concept of Ability	
How good are you in this subject?	*Not good at all; very good*	0.87
How good do you think you are in this subject compared to other students in the class?	*Much worse than other students; much better than other students*	0.88
How often do you feel smart in this class?	*Never; very often*	0.68
How much natural ability do you have in this subject?	*No ability at all; a lot of ability*	0.79
	Change in Effort	
If a student works to his or her highest potential in a class, then we could say that he or she is putting forth 100% effort to learn the subject matter. How much effort do you usually put forth in this class?	*I am not trying at all* (0%); *I am working to my highest potential* (100%)	11.82
How hard are you working to learn about this subject?	*Not hard at all; as hard as I can*	0.96
How hard do you study for tests in this class?	*Just enough to pass; whatever it takes to get a good grade*	0.61
How hard do you work in this class?	*Much less than most classes; much more than most classes*	0.72
If a student works to his or her highest potential in a class, then we could say that he or she is putting forth 100% effort to learn the subject matter. Please estimate how much effort each student listed below is putting forth in this class. [From teacher's questionnaire]	*Student is not trying at all* (0%); *student is working hard enough to fulfill his or her highest potential* (100%)	4.30

Note. All items are from the student's questionnaire, with the exception of the final item. Response scales for each item range from 1 to 7, with the exception of the first and last items in the Change in Effort factor, for which responses were rated on an 11-point scale, ranging from 0% to 100%. Factor loadings are derived from the standardized solution, in which the factors (but not the measured variables) have been standardized. Thus, each loading indicates the expected change in the raw score of a measured variable given an increase of 1 standard score in the factor. The raw difference scores for the first and last item in the Change in Effort factor have a possible range of −100 to +100; the possible range for every other measured variable is −6 to +6.

Received July 5, 1990
Revision received October 18, 1990
Accepted December 6, 1990 ■

© International Reading Association 0034-0553/93/US$1.25+.00

Barbara A. Wasik

Robert E. Slavin
Center for Research on Effective Schooling for Disadvantaged Students
The Johns Hopkins University, Baltimore, Maryland

Preventing early reading failure with one-to-one tutoring: A review of five programs

Tutoring is the oldest form of instruction. Parents have always provided one-to-one instruction to their children, and learning settings from driving instruction to on-the-job training typically employ one teacher for each learner for at least part of the learner's instruction.

In elementary and secondary instruction, one-to-one tutoring exists around the margins of group instruction. For example, teachers often work with individual children during seat work periods, recess, study hall, or after school. Parents often hire tutors to work with their children. Tutoring is often used in special education, and sometimes in other remedial programs such as compensatory education.

The topic of tutoring has come to the fore in recent years because of a renewed focus on students who are at risk of school failure, coupled with a renewed commitment to see that all students learn basic skills in the early grades. In particular, modest effects of traditional U.S. Chapter 1/Title I pullout programs (Carter, 1984) and the loosening of restrictions on uses of Chapter 1 funds have contributed to a broader range of services being provided under Chapter 1 funding.

One-to-one tutoring is one option often being considered or implemented. In recent years, increased flexibility in Chapter 1 and other factors have led to the use of tutors with first graders to prevent early reading failure. Advocates of tutoring programs argue that first grade is a critical year for the learning of reading, and reading success in the early grades is an essential basis for success in the later grades. Clay (1979), for example, argues that early intervention for children who have problems learning to read is crucial to children's later success. For students who do not learn to read in traditional classrooms or with traditional reading programs, one-to-one tutoring is a possible solution to preventing early reading failure.

Research on Chapter 1 programs suggests that remediation of learning problems after the primary grades is largely ineffective (see Kennedy, Birman, & Demaline, 1986). It may be that it is easier to prevent learning problems in the first place than to attempt to remediate them in the later grades. Considering how much progress the average reader makes in reading between the first and last days of first grade, it is easy to see how students who fail to learn to read during first grade are far behind their peers and will have difficulty catching up.

The major drawback to tutoring is its cost. Providing tutoring to large numbers of students across the grade span would, of course, be prohibitive. But if in fact early intervention can prevent children from experiencing failure and can help them get off to a successful start in school, the use of this expensive intervention may be cost effective in the long run.

The importance of understanding the effects of first-grade tutoring goes far beyond the pedagogical and

Reprinted with permission of Barbara A. Wasik and the International Reading Association

technical issues involved. Edmonds's (1981) statement that every child can learn and Bloom's (1981) assertion to the same effect contributed to a variety of discussions among policy makers about learning as an "entitlement" for all children, on the basis that if every child can learn, then schools have an ethical and perhaps legal responsibility to see that every child does learn. One manifestation of this point of view is a document produced by the Council of Chief State School Officers (1987) that describes model state statutes to entitle every U.S. child not only to an appropriate education but to success in achieving an acceptable level of performance (also see Council of Chief State School Officers, 1989). If success is seen as an entitlement, educators must have methods that produce success for all nonretarded children regardless of home background, no matter how expensive these methods may be. In any discussion along these lines, one-to-one tutoring for at-risk students is sure to be one element of the strategy to ensure success for all.

Recently, there is an unprecedented willingness among educators to adopt expensive early intervention programs if they are believed to reliably produce large effects. Examples of this include Project STAR in Tennessee (Word et al., 1990) and Project Prime Time in Indiana (Farr, Quilling, Bessel, & Johnson, 1987), which have implemented substantially reduced class sizes in the early elementary grades. Growing provision of preschool and extended day kindergarten programs and of IBM's Writing to Read computer program are other examples. Recently, many districts have adopted Reading Recovery and Success for All, intensive reading programs with tutoring, as means of preventing early school failure.

It would be important to know the effectiveness of such programs that are expensive to implement and maintain in school districts. If school districts plan to allocate Chapter 1 funds to expensive programs, the effectiveness of these programs should be of great concern. It is important to know how large the effect of tutoring is (in comparison to plausible alternatives), to what degree effects of tutoring are maintained over time, and which specific tutoring programs and practices produce the largest gains in student reading achievement. The purpose of this article is to review the research on the effectiveness of one-to-one tutoring programs to identify what is currently known about the answers to these and other questions.

Previous reviews of research on tutoring have primarily focused on peer tutoring (e.g., Devin-Sheehan, Feldman, & Allen, 1976; Scruggs & Richter, 1985). The one review that included tutoring by adults primarily focused on applications in special education (Polloway, Cronin, & Patton, 1986). None of these earlier reviews discussed any of the first-grade reading prevention models emphasized here.

In the present article, we consider the effectiveness of tutorial programs from two perspectives: empirical and pragmatic. From the empirical perspective, one can ask questions such as "Does the program work?" and "How strong are its effects?" To answer these questions, we computed effect sizes for each of the five programs. (This is discussed in detail in the section on review methods.) From a pragmatic standpoint, one can ask questions such as "What components of reading are included?" and "Does it matter if the tutors are certified teachers or paraprofessionals?" and "Why are some programs apparently more effective than others?"

It would also be important to examine the theoretical similarities and differences of these programs regarding the approach taken to learning in general, and reading in specific, and how the relationship between the tutor/student dyad facilitates learning. One aspect of effectiveness of tutorial programs could be explained by appealing to domain-general theories such as Vygotsky (1978) that have been formulated to account for the transmission of knowledge in one-to-one dyads. However, while the Vygotskian perspective has been explored with one program, Reading Recovery (see Clay & Cazden, 1990), theories to account for transmission of knowledge from tutor to student have not been explored in the other programs. Similarly, it would be important to examine the different theories of reading as espoused by advocates of each tutoring program. However, again with the exception of Reading Recovery, the programs do not articulate a theory of reading.

In what follows, we review five tutoring programs. In the course of describing these programs, we discuss the model of reading to which each program subscribes to and identify the key components of reading found in each program. From reviewing the curriculum of the tutoring programs, we have identified eight components of the reading process that are emphasized in these programs: perceptual analysis of print, knowledge of print conventions, decoding, oral language proficiency, prior knowledge, lexical access, syntactic analysis of sentences, and prose comprehension. We acknowledge that this is by no means a complete list, since key aspects of reading such as phonemic awareness are not included. However, these components were extracted from the programs reviewed. We then discuss which components each program includes. We also consider the nature of the tutors and how the programs are implemented. Then we provide effect sizes to qualify the empirical effects of the programs. If one tutoring program appears to be more effective than another, it could be because (a) practical differences in the program lead to different out-

80

comes, (e.g., certified teachers are used in one and not the other), or (b) tutors in one are using more effective methods or curricula than those in the other, or (c) different programs to emphasize to different degrees or reading components that are considered to be central in contemporary theories of reading. In our discussion, we consider these and other explanations.

Review methods

This review uses a set of procedures called best-evidence synthesis, which combines elements of meta-analysis with those of traditional narrative reviews (Slavin, 1986). Briefly, a best-evidence synthesis requires locating all research on a given topic and discussing the substantive and methodological issues in the research as in a narrative review. A prior criteria for germaneness to the topic at hand and for methodological adequacy are typically applied. Whenever possible, study outcomes are characterized in terms of effect size (ES), the difference between experimental and control means divided by the control group standard deviation. When means or standard deviations are not reported, effect sizes are estimated from *F*, *t*, or other statistics (see Glass, McGaw, & Smith, 1981). The numerator of the effect size formula may be adjusted for pretests or covariates by computation of gain scores or use of ANCOVA, but the denominator is always the unadjusted individual level standard deviation of the control group or (if necessary) a pooled standard deviation.

Inclusion criteria. Studies were included in the present review if they evaluated one-to-one instruction delivered by adults (certified teachers, paraprofessionals, or volunteers) to students in the first grade who are learning to read for the first time. Studies had to compare tutoring to traditional instruction in elementary schools over periods of at least 4 weeks on measures of objectives pursued equally in experimental and control conditions. This duration requirement did not exclude any studies of first-grade tutoring. The first-grade requirement excluded only three studies (Bausell, Moody, & Walzl, 1972; Fresko & Eisenberg, 1983; and Shaver & Nuhn, 1971), which looked at remedial tutoring in the third grade and higher. Studies of cross-age and same-age peer tutoring (e.g., Cloward, 1967; Greenwood, Delquardi, & Hall, 1989; and von Harrison & Gottfredson, 1986), did not fit this criterion and were not included. All studies ever written in English were included. The only study done outside of the U.S., by Clay (1985), examined only students who were successful in tutoring, not all who received it. This study is described in the section on Reading Recovery. Therefore, this best-evidence synthesis included all methodologically adequate studies of one-to-one tutoring that focused on instruction delivered by adults to first graders. In a complete review of published as well as unpublished studies, a total of 16 studies met the inclusion criteria.

Research on preventive tutoring programs

All of the studies that met the inclusion criteria specified above evaluated a total of five tutoring programs. These programs incorporated instructional materials as well as provision of one-to-one tutors. Some of the major characteristics of these programs are summarized in Table 1. Table 2 provides additional detail on models of reading used in each program. As is apparent from the Tables, the five programs vary widely in curriculum, integration with classroom instruction, use of certified versus paraprofessional tutors, and other factors not intrinsically related to the one-to-one setting. These programs also differ in their model of reading and the measures used to assess the effectiveness of these programs. As a result, we make no attempt to combine findings across studies in any way. However, we do discuss how different approaches to reading translate into the method used in the tutoring process. Finally, we discuss how ultimately the reading model is tied to the type of assessment each program uses to evaluate its effectiveness, suggesting that curriculum, instruction, and assessment are interrelated (Weade, 1987).

Reading Recovery

The preventive tutoring program that has received the most attention and use in recent years is Reading Recovery. This program was originally developed by Marie Clay (1985) in New Zealand, and is widely used in that country. In 1984-85, Marie Clay and a colleague, Barbara Watson, spent a year at the Ohio State University. They trained a group of teachers to use the program, and trained several Ohio State faculty members to train others. Since that time, research on Reading Recovery has been conducted at Ohio State, and the program has rapidly expanded in use.

As applied in the longitudinal studies, Reading Recovery provides one-to-one tutoring to first graders who score in the lowest 20% of their classes on a program-developed diagnostic survey. The tutors are certified teachers who receive training for 2.5 hours per week for an entire academic year. Students are tutored for 30 minutes each day until one of two things happen. If students reach the level of performance of their classmates in the middle reading group, they are "discontinued." If they receive 60 lessons without achieving this

READING RESEARCH QUARTERLY April/May/June 1993 28/2

Table 1 Characteristics of preventive tutoring programs

Program	Location of evaluations	Tutors	Tutees	Duration	Tutoring methods and curriculum
Reading Recovery	Ohio; Chicago, Illinois	Certified reading teachers	Lowest first graders	30 minutes/day ranging from 12 to 20 weeks	Learning to read by reading. Reading short stories and connecting writing activities to reading. Tutors guide children to learn metacognitive strategies. No connection to classroom instruction.
Success for All	Inner-city Baltimore, Maryland	Certified teachers	Lowest first and second graders	20 minutes/day evaluted on 8-week cycle	Learning to read by reading. Closely integrated with structured classroom curriculum. Emphasis on metacognitive strategies.
Prevention of Learning Disabilities	New York; Ohio; California	Certified teachers	Lowest first and second graders	30 minutes, 3 to 5 times/week	Use of directed activities to teach specific perceptual and spatial skills involved in reading. Emphasis on skill acquisition. No emphasis on reading connected text. No connection with a classroom curriculum.
Wallach Tutoring Program	Inner-city Chicago, Illinois; rural North Carolina	Paraprofessionals	Lowest first graders	30 minutes/day, 1 year	Phonics-based tutoring program. Emphasis on systematic mastery of phonetic skills. Does not focus on reading connected text. Not integrated with classroom instruction.
Programmed Tutorial Reading materials; includes	Inner-city Indianapolis, Indiana; Lenoir City, North Carolina	Paraprofessionals	All first graders	15 and 30	Highly detailed and prescribed lessons minutes/day with corresponding sight-reading program, comprehension, and word analysis. Emphasis on skills. Partially integrated with classroom instruction.

Table 2 Components of reading emphasized in tutoring programs

	Programs				
Components	Reading Recovery	Success for All	Prevention of Learning Disabilities	Wallach Tutoring Program	Programmed Tutorial Reading
Perceptual analysis of print	Yes	Yes	Yes	Yes	Yes
Knowledge of print conventions	Yes	No	Yes	No	No
Decoding	Yes	Yes	Yes	Yes	Yes
Oral language proficiency	Yes	Yes	Yes	Yes	No
Prior knowledge	Yes	Yes	No	No	No
Lexical access	No	No	No	No	No
Syntactic analysis of sentences	No	No	No	No	No
Prose comprehension					
Prose structure	No	No	No	No	No
Story grammar	No	No	No	No	No
Inference making	Yes	Yes	Yes	No	No
Reading strategies	Yes	Yes	No	No	No
Metacognition and error detection	Yes	Yes	No	No	No
Error correction strategies	Yes	Yes	No	No	No

level of performance, the students are released from the program but considered "not discontinued."

Model of reading

In the Reading Recovery program, reading is viewed as a psycholinguistic process in which the reader constructs meaning from print (Clay, 1979; Pinnell, 1985). According to Clay, reading is defined as a "message-gaining, problem solving activity, which increases in power and flexibility the more it is practiced." Clay states that within the "directional constraints of the printer's code, language and visual perception responses are purposefully directed in some integrated way to the problem of extracting meaning from text, in sequence, to yield a meaningful communication, conveying the author's message" (Clay, 1979, p. 6).

Clay does not specifically address how language and visual perception are coordinated in order to extract meaning from text. Nevertheless, her discussion of reading and components of the Reading Recovery Program suggest that she includes the following components of reading in her model: perceptual analysis, knowledge of print conventions, decoding, oral language proficiency, prior knowledge, inference making, reading strategies, metacognition and error detection, and error correction strategies (see Table 2).

Clay (1979) describes reading as the "process by which the child can, on the run, extract a sequence of cues from printed texts and relate these one to another, so that he can understand the precise message of the text." In order to master this process, the child must have good control of oral language, developed perceptual skills, the physiological maturity and experiences that allow the child to coordinate what s/he hears in language and sees in print, and enough hand-eye coordination so s/he can learn the controlled, directional patterns required for reading (Clay, 1979). Expert teachers are assumed to have sufficient implicit knowledge of the processes that they can recognize the source of the child's difficulty.

From this theory of reading, three major theoretical principles serve as a foundation for the Reading Recovery Program. First, reading is considered a strategic process that takes place in the child's mind. Reading requires the coordination of many strategies and visual information, the integrating of letter-sound relationships, features of print, and the child's own background knowledge. Meaning is never derived just from the print alone, but from the interaction of the reader's unique background and the print. Second, reading and writing are interconnected. Having the child make the connection between reading and writing is essential to literacy development. Third, "children learn to read by reading" (Pinnell, 1989; Pinnell, DeFord, & Lyons, 1988). Children must engage in reading of connected text and should avoid working on isolated skills in order to become proficient in reading. It is only by reading frequently that the child can come to detect regularities and redundancies present in written language.

These three principles set the foundation for the Reading Recovery program. Children in Reading Recovery spend most of their time engaged in reading and writing activities. There is no systematic presentation of phonics, yet during the reading and writing activities, letter-sound relationships are taught as one of the basic strategies in solving problems. Tutors use a variety of strategies to help students develop "independent, self-generating systems for promoting their own literacy" (Pinnell, 1985).

Structure of tutoring

For the first few tutoring sessions, the teacher and student "roam around the known," reading and writing together in an unstructured, supportive fashion, to build a positive relationship and to give the teacher a broader knowledge of the child. After this, teachers begin to use a structured sequence of activities, as follows (adapted from Pinnell et al., 1988, pp. 10-11).

The child rereads familiar books. The child reads again several favorite books that s/he has previously read. The materials are storybooks with natural language rather than controlled vocabulary. Books within a lesson may range from quite easy to more challenging, but the child is generally reading above 90% accuracy. During this time, the child has a chance to gain experience in fluent reading and in using strategies "on the run" while focusing on the meaning of the text. The teacher interacts with the child during and after the reading, not "correcting," but talking with the child about the story and supporting the effective actions the child has taken.

The teacher analyzes reading using the running record. Each day the teacher takes a running record of a book that was new for the child the previous day. The running record is a procedure similar to miscue analysis (Goodman, Watson, & Burke, 1987). Using a kind of shorthand of checks and other symbols, the teacher records the child's reading behavior during oral reading of the day's selected book. The teacher examines running records closely, analyzing errors and paying particular attention to behavior such as self-correction. In this way, s/he determines the strategies the child is using to gain meaning from text. This assessment provides an ongoing picture of the progress the child makes. While the child is reading, the teacher acts as a neutral observer; the child works independently. The accuracy check tells the teacher whether the text was well selected and

Table 3 First-year evaluations of Reading Recovery

Measure	Effect sizes Pilot cohort	Effect sizes Second cohort
Letter identification	+.36	-.04
Word test	-.13	+.40
Concepts about print	+.60	+.65
Writing vocabulary	+.62	+.69
Dictation	+.57	+1.03
Text reading	+.72	+.91

Note. Pilot cohort data are from Huck & Pinnell, 1986; second cohort data are from Pinnell, Short, Lyons, & Young, 1986. There are apparent ceiling effects on the letter identification and word tests.

introduced the day before.

The child writes messages and stories and then reads them. Every day the child is invited to compose a message and to write it with the support of the teacher. Writing is considered an integral part of gaining control over messages in print. The process gives the child a chance to closely examine the details of written language in a message that s/he has composed, supported by her/his own language and sense of meaning. Through writing, the child also develops strategies for hearing sounds in words and using visual information to monitor and check her/his reading.

After the construction of the message, the teacher writes it on a sentence strip and cuts it up for the child to reassemble and read. This activity provides a chance to search, check, and notice visual information. Using plastic letters on a magnetic board, the teacher may take the opportunity to work briefly with the letters to increase the child's familiarity with the names of letters and their use in known words such as the child's name. This work will vary according to the knowledge the child already has.

The child reads new books. Every day the child is introduced to a new book that s/he will be expected to read without help the next day. Before reading, the teacher talks with the child about the book as they look at the pictures. The teacher helps the child build a frame of meaning prior to reading the text. The purpose of the introduction is not necessarily to introduce new words, but to create understanding in advance of reading so that it will be easier to focus on meaning.

Every child's program differs. Children do a great deal of reading, but not from a graded sequence. No child reads the same series of books. The small books are carefully selected by the teacher for that child at that time. In writing, children work on their own messages, so they are writing and reading works that are important to them individually. The major difference within and across lessons lies in the teacher's ability to follow each child and to respond in ways that support acceleration and the development of strategies. Strategies may include directional movement, one-to-one matching, self-monitoring, cross-checking, using multiple cue sources, and self-correction. The Reading Recovery teacher uses instructional techniques designed to help the child develop and use such strategies.

The tutoring model in Reading Recovery is separate from the instruction provided in the regular classroom. Most often, Reading Recovery teachers tutor students half time and either teach small groups of Chapter 1 students or teach a regular class the other half. The tutees may thus have the same teacher as their reading teacher and as their tutor, but in general this does not occur.

Tutor training in Reading Recovery is extensive. During the first year, in addition to teaching a reading class and tutoring four students, the tutors attend weekly seminars during which they receive training in observational, diagnostic, and assessment techniques and are schooled in the reading philosophy of Marie Clay. The tutors also participate in weekly "behind the glass" demonstration lessons where they observe actual tutoring sessions behind a one-way mirror and have the opportunity to critique and discuss the lesson. Considerable time is spent learning about the reading process and learning how to implement appropriate strategies to meet the needs of individual children. Follow-up inservice training continues after the first year. Additional training is required of Teacher Leaders who are certified to train Reading Recovery tutors in their areas. Teacher Leaders participate in a 1-year internship at the Ohio State University training center (other states such as New York are establishing regional centers), where they participate in reading and writing seminars and learn to train tutors using the "behind the glass" technique.

Results

Research evaluating Reading Recovery in New Zealand (Clay, 1985) focused entirely on the discontinued students (those who were successful in the program), and therefore does not provide a full account of the effectiveness of the intervention. However, the U.S. research has included discontinued and not discontinued students—all of the students who either graduated from the program or received at least 60 lessons.

Preventing early reading failure

The Ohio State group has conducted two longitudinal studies comparing Reading Recovery to traditional Chapter 1 pullout or in-class methods. The first (pilot) study (Huck & Pinnell, 1986; Pinnell, 1988) of Reading Recovery involved 21 teachers trained by Marie Clay who worked in six inner-city Columbus, Ohio, schools. Each school provided one Reading Recovery class and a matched comparison class. The lowest 20% of students in each class served as the experimental and control group, respectively. Students were pretested in September and December, 1984, but the tutoring did not begin until the spring semester, 1985.

The second longitudinal study (DeFord, Pinnell, Lyons, & Young, 1988; Pinnell, Short, Lyons, & Young, 1986) involved 32 teachers in 12 schools in Columbus. Twelve of these teachers had been tutors in the pilot cohort. In this study, students in the lowest 20% of their classes were randomly assigned to Reading Recovery or control conditions. The research design originally made a distinction between students in the experimental and control groups who had Reading-Recovery-trained versus non-Reading-Recovery-trained teachers in their regular reading program. However, there were no differences on this factor, so the analyses focused on tutored versus untutored children, regardless of who their regular reading teacher was.

The results at the end of the first implementation year for the two Ohio State studies are summarized in Table 3. Reading Recovery students substantially outperformed control students on almost all measures. The exceptions were tests of letter identification and a word recognition scale, which had apparent ceiling effects in both conditions.

Each spring for 2 years following the implementation year, all children were assessed on Text Reading Level, an individually administered test in which students are asked to read from books with progressively more difficult content. This measure yields a reading level (e.g., second grade, first semester).

The results on this measure, summarized in Table 4, show an interesting statistical paradox. By the criterion of effect size, the effects of Reading Recovery are clearly diminishing each year. By the end of the third grade, the effect size for the pilot cohort has diminished from +.72 to +.14, and in the second cohort the effect size diminished from +.78 to +.25. On the other hand, the difference in raw units between Reading Recovery and control students remained about the same across all 3 years, hovering around two points in the pilot cohort and three in the second cohort. Is the effect maintaining or not?

The difference between these two measures is that the standard deviation of the Text Reading Level measure increases each year, making the same raw difference a smaller proportion of the standard deviation. In more substantive terms, the size of the difference may not be diminishing (assuming the measure is an equal-interval scale), but the importance of the difference is diminishing. For example, a difference of 3 months on a standardized reading test might be a big difference at the end of the first grade but is a small one at the end of sixth grade.

Table 4 Longitudinal evaluations of Reading Recovery

Time of evaluation	Effect sizes (raw differences) Pilot cohort	Second cohort
End of implementation year	+.72 (1.6)	+.78 (2.8)
1-year follow-up (Grade 2)	+.29 (2.0)	+.46 (3.0)
2-year follow-up (Grade 3)	+.14 (1.8)	+.25 (2.8)

Note. All data are from individually administered text reading level assessments developed by the program developers. Pilot cohort data are from Pinnell, 1988; second cohort data are from DeFord, Pinnell, Lyons, & Young, 1988.

Actually, there is a more complex story on the longitudinal effects of Reading Recovery. The students who succeeded in Reading Recovery, those categorized as discontinued, were performing on average at a level like that of their classes as a whole, and substantially better than the comparison group of low achievers. On the other hand, all of the not-discontinued students (who had at least 60 tutoring sessions but failed to achieve at the level of the rest of their class) were still below the level of their classmates by third grade, and were substantially lower than the control group. These not-discontinued students represented 27% of the former Reading Recovery students tested in the third grade in the second cohort study (DeFord et al., 1988).

Effects of Reading Recovery on promotions from grade to grade. Participation in Reading Recovery increased students' chances of being promoted to the second grade in comparison to the control low achievers. Although 31% of comparison students were retained in first grade or assigned to special education, this happened to only 22% of Reading Recovery students (DeFord et al., 1988). However, by the third grade this difference had mostly disappeared. Two years after the children were in the first grade, a total of 59.6% of Reading Recovery children and 57.8% of control children were in the third grade 2 years after first grade. A school district evaluation in Wakeman, Ohio, found that first-grade retentions dropped from 24 to 1 in the 3 years

Table 5 Ohio statewide study of Reading Recovery (adjusted effect sizes in comparison to control groups)

Measure	Reading Recovery	Reading Success	DISP Tutoring	Reading-Writing Group
February				
Dictation	+.65	+.45	-.05	+.14
Text Reading Level	+1.50	+.45	-.01	+.41
Woodcock	+.49	+.04	+.25	+.23
Gates	+.51	+.27	+.14	+.23
May				
Gates-MacGinitie	+.19	-.14	-.05	+.34
October				
Dictation	+.35	+.00	-.25	+.29
Text Reading Level	+.75	+.07	+.06	+.32

Adapted from Pinnell, Lyons, DeFord, Bryk, & Seltzer, 1991

after implementation of Reading Recovery (Lyons, Pinnell, Deford, McCarrier, & Schnug, 1989).

One additional study compared Reading Recovery to control treatments in first grade. This was a study conducted in four Chicago elementary schools. As in the earlier studies, students were randomly assigned to Reading Recovery or control conditions. Because neither standard deviations nor statistical tests are presented, effect sizes cannot be computed, but program effects in comparison to control students were clearly substantial. Applying standard deviations from the Ohio studies to the same measures used in Chicago yields end-of-first-grade effect sizes of approximately +.90 on dictation and text reading level.

The most recent major study of Reading Recovery conducted by the Ohio State group (Pinnell, Lyons, DeFord, Bryk, & Seltzer, 1991) evaluated the full program in comparison to three alternative programs and a control group in 10 Ohio school districts. The treatments were as follows:

1. *Reading Recovery* (RR) was implemented as in earlier assessments.
2. *Reading Success* (RS) was the same as Reading Recovery except that teachers received a 2-week training session in the summer instead of the yearlong, 2 to 3 hours per week training with "behind the glass" demonstration teaching used in Reading Recovery. In comparison to Reading Recovery, this treatment tested the possibility that effects like those for the program as usually implemented could be obtained with far less extensive training, a major stumbling block to widespread diffusion.
3. *Direct Instruction Skills Plan* (DISP) was an individual tutorial program that tested the possibility that the one-to-one tutoring, not the particulars of the Reading Recovery model, explains the effects of the program. DISP used direct instruction in specific skills such as letter, sound, and word recognition, sequencing, filling in blanks, answering questions, and reading extended text. Teachers were encouraged to design lessons themselves to teach these and other skills.
4. *Reading and Writing Group* (RWG) was a small group tutorial model taught by teachers who had been trained as Reading Recovery teachers. They used Reading Recovery materials and strategies but were asked to adapt them to the small group setting in their own ways. This treatment essentially tested the effects of the one-to-one tutoring aspect of Reading Recovery, holding curriculum constant.
5. *Control group* for each treatment was the Chapter 1 pullout program already in existence in each school.

Four schools (one per treatment) were involved in each district. In each school that already had a Reading Recovery teacher, students were randomly assigned to RR or control (Chapter 1) treatments. In other schools, additional teachers were hired from the district's substitute lists to implement the RS or DISP tutoring models. Trained Reading Recovery teachers were added to schools to implement the Reading and Writing Group (RWG) treatment. Students were randomly assigned to treatment or control classes.

The treatments were implemented starting early in first grade. Students were then assessed in February, in May, and again in the following October. The results are summarized in Table 5.

As is clear from Table 5, the effects varied considerably according to measure and time of test administration.[1] The February measures clearly favored the Reading Recovery on all measures and the Reading Success model on the two measures developed as part of the Reading Recovery program, Dictation and Text Reading Level. However, the February measures are biased in favor of the three tutoring models. By February, the tutoring was concluded, and students moved into the Chapter 1 group program. In contrast, the RWG and Chapter 1 control group programs were yearlong interventions, so measuring effects in February discriminates against them.

Unfortunately, the only test given in May was the standardized Gates-MacGinitie, which found few effects for any treatment.

The October follow up provides the best indication of the effects of the four programs. The most positive effects were found for Reading Recovery on Dictation (ES = +.35) and Text Reading Level (ES = +.75). Neither of the other two tutoring methods (RS and DISP) found any positive effects. It is interesting to note that after the full program, it was the Reading and Writing Group (RWG) treatment that had the most positive effects (ES = +.29 for Dictation, +.32 for Text Reading Level). This treatment also had the largest positive effects on the May Gates-MacGinitie of all treatments (ES = +.34).

One important factor may be confounded with the effects of the four programs. The teachers in the two most successful treatments, Reading Recovery and Reading and Writing Group, were experienced Reading Recovery teachers who had a year of Reading Recovery training and at least a year of experience in implementing the program. In contrast, the Reading Success and DISP teachers were hired from the substitute list and may have been considerably less skilled and less experienced.

At a minimum, the Ohio statewide study provides one more convincing evaluation of Reading Recovery, showing large effects, especially on Text Reading Level, which maintain into the school year following the intervention. The findings suggest that the yearlong training, the particular curriculum and instructional model used, and the one-to-one aspect of the tutoring are all critical to the success of the model, but these conclusions may be tempered by possible differences in teacher quality in the groups that received shorter training (RS) and the alternative tutoring model (DISP).

A few methodological issues about the Reading Recovery research are worth raising. First, there is an articulation between the Reading Recovery program and the measures used to evaluate the program, suggesting that what is taught is what is measured. The measures used were all individually administered scales designed either by Marie Clay and her associates or by the Ohio State researchers. Five of the measures, Letter Identification, Word Test, Concepts about Print, Writing Vocabulary, and Dictation, make up the Diagnostic Survey, which was developed by Clay. The Letter Identification test asks students to identify 54 letters in upper and lower case. The Word Test is a list of high-frequency words from the basal reader used in the school district. Concepts about Print asks the students to identify conventions of print and reading. The Writing Vocabulary has the children write down as many words as they can, starting with their own name, in 10 minutes. The Dictation test assesses children's ability to write down every word in a sentence that is read to them. In scoring this test, children are given credit for every sound correctly represented. The Text Reading Level is the sixth test administered in the Reading Recovery evaluation. This test consists of a series of graded stories that the child reads. A running record of the child's oral reading is taken and then an accuracy level is calculated. These measures correspond to the model of reading in Reading Recovery. As discussed earlier, the reading model emphasizes oral language, perceptual analysis, concepts of print, reading strategies, and metacognition. All of these aspects are emphasized in the outcome measures. Therefore, children who were tutored in Reading Recovery were also more familiar with the assessment than were the children in the control groups.

It also appears that bias in favor of the kinds of skills taught in the program is most likely at the low levels of the Text Reading Level measure, where assessments focus on concepts of print, using pictures and patterns to guess story content, and other skills specifically taught in Reading Recovery. The finding of particularly large effects on Text Reading Level (in contrast to other measures) was especially pronounced in the Ohio statewide study (Pinnell et al., 1991).

Secondly, Reading Recovery has a policy of not serving students who have already been retained in first grade and students identified for special education. One of the reports (Pinnell et al., 1986) implies that some students originally selected for tutoring failed to make adequate progress in early tutoring sessions and were excused from tutoring (and therefore excluded from the evaluation). Any of these practices might have influenced the Reading Recovery sample by excluding the very lowest achievers.

These criticisms aside, the effects of Reading Recovery are impressive at the end of the implementation year, and the effects are maintained for at least 2 years. In addition, the Ohio State researchers have studied implementation issues that affect the quality of the program. For example, Lyons (1991) studied the effects of duration of training on Reading Recovery teachers. Teachers who had a 2-week inservice program were compared to teachers who attended a yearlong training program. The results showed that students who had teachers who received more extensive training outperformed students who had teachers in the 2-week program on Text Reading Level.

In another study, Handerhan (1990) conducted a sociolinguistic analysis of teachers and children in Reading Recovery. Reading Recovery tutoring sessions were videotaped and sessions of four of the most and least successful teachers (based on what was accomplished with the student) were analyzed. Handerhan (1990) found that across tutors there was consistency in how they structured the lessons regarding similarities in language, materials, and procedural techniques. However, more successful tutors showed greater variability in the strategies they used and the less successful tutors engaged more in presenting letters and words as discrete skills without reading for meaning. This study is important because it documents the variability in instruction during tutoring as well as identifying what behaviors are necessary to be a successful tutor helping children learn to read. The rapidly expanding use of Reading Recovery throughout the U.S. (see Lyons, Pinnell, DeFord, McCarrier, & Schnug, 1989) shows that the program is practical to use.

Success for All

Success for All (Madden et al., 1991; Slavin, Madden, Karweit, Dolan, & Wasik, 1992; Slavin, Madden, Karweit, Livermon, & Dolan, 1990) is a comprehensive schoolwide restructuring program that is designed primarily for schools serving large numbers of disadvantaged students. Its main intention is to see that all children are successful in basic skills, particularly reading, the first time they are taught. One major element of Success for All is one-to-one tutoring by certified teachers for students in Grades 1-3 who are having difficulties learning to read. The program includes many other elements, such as a beginning reading program, preschool and kindergarten programs, and family support services. However, for low-achieving first graders, who receive most of the tutoring services, the Success for All program can be seen primarily as a preventive tutoring program.

Model of reading

The Success for All tutoring program is based on research that "points to the need to have students learn to read in meaningful contexts and at the same time have a systematic presentation of word attack skills" (Slavin et al., 1992). Its underlying philosophy is that there is certain regularity to language, and that direct presentation of phonics is viewed as a helpful strategy which children can use to figure out words. Children also need to build a strong sight vocabulary that will help in identifying words that are not decodable. Along with the systematic presentation of phonics, children engage in reading meaningful connected text. The Success for All program emphasizes that reading is a strategic process that takes place in the student's mind and that these strategies should be taught directly.

Unlike Reading Recovery, Success for All does not articulate a complete theory of beginning reading. However, an underlying model of reading can be seen in the structure and content of the program. There are four components that drive the Success for All tutoring program. First, children learn to read by reading meaningful text. Reading skills are not acquired by children learning isolated, unconnected information about print. Second, phonics needs to be taught systematically as a strategy for cracking the reading code. Children engage in reading stories that are meaningful and interesting, yet have a phonetically controlled vocabulary. Third, children need to be taught the relationship between reading words and comprehending what they read. Mere word recognition is not reading. The emphasis on comprehension is directly related to the fourth component, the emphasis on children's need to be taught strategies to help them become successful readers. Children who have problems learning to read often do not know how to effectively use metacognitive strategies to help them read. Through direct instruction, children are taught when, how, and why they should use strategies.

In summary, Success for All emphasizes the following components in its model of reading: perceptual analysis, decoding, prior knowledge, oral language proficiency, inference making, reading strategies, metacognition and error detection, and error correction strategies.

Structure of tutoring

The tutoring model used in Success for All is different in many ways from that used in Reading Recovery. One difference is that in Success for All, the tutoring model is completely integrated with the reading program. The tutor's most important responsibility is to make sure that the student is making adequate progress on the specific skills and concepts being taught in the reading class.

Another difference is that in Success for All, first graders receive tutoring as long as they need it. Although most students receive tutoring for part of a year, some receive it all year and then continue to be tutored into the second grade. The commitment in Success for All is to see that every child succeeds, that no child is retained or assigned to special education except under extreme circumstances.

First graders are initially selected into tutoring in Success for All on the basis of individually administered informal reading inventories given in September. After that, however, students are assessed every 8 weeks in terms of their progress through the reading curriculum. On the basis of these 8-week assessments, students who are doing well may be rotated out of tutoring as other students are rotated into tutoring. The amount of tutoring received by a given student may vary from 8 weeks to the entire year or more.

Students receive tutoring every day for 20 minutes. This time is usually scheduled during an hourlong social studies/science block, so that tutoring represents additional time in reading.

The tutors are certified teachers recruited in the same way as other teachers. Each tutor teaches a 90-minute reading class each day (to reduce class size for reading) and then spends the rest of the day tutoring three children per hour. Because the tutors teach a reading class, they are fully aware of what the reading program is; if a child is struggling with Lesson 37, the tutor knows exactly what is required for success in Lesson 37 because he or she has taught it.

In many cases, tutors work with students who are also in their morning reading class. When scheduling does not allow this, the student's reading teacher fills out a tutor/teacher communication form that indicates what lesson the student is working on in class and the teacher's assessment of the specific problems the student is having with that lesson. The tutor uses this information to plan the tutoring session. This communication ensures coordination between the classroom instruction and tutoring.

The tutors receive 2 days of training (along with all other beginning reading teachers) to learn to teach the Success for All beginning reading program (described below), and then they receive 4 additional days of training on assessment and on tutoring itself. Tutors are observed weekly by the program facilitator and given direct feedback on the sessions.

A strong emphasis is placed on teaching comprehension strategies. The tutor's goal is to get the students to read fluently, and to understand what they read. Tutors are trained to explicitly teach metacognitive strategies to help students monitor their comprehension. For example, a tutor will teach a student to stop at the end of each page and ask, "Did I understand what I just read?" The students learn to check their own comprehension and to go back and reread what they did not understand.

Each tutoring session is structured, but the tutor is constantly diagnosing and assessing the individual needs of each student and tailoring the sessions to fit the student's specific problem. If a student is having difficulty with fluency, the tutor will have the student do repeated reading aloud of a story. With similar materials, a tutor may work with another child on comprehension monitoring.

A typical tutoring session begins with the student reading out loud a familiar story that he or she has read before in tutoring and in the reading class. This is followed by a 1-minute drill of letter sounds to give the student the opportunity to practice the letter sounds taught in class. The major portion of the tutoring session is spent on reading aloud "shared stories" that correspond to the beginning reading lessons. The shared stories are interesting, predictable stories that have phonemically controlled vocabulary in large type and other elements of the story in small type. The teacher reads aloud the small-type sections to provide a context for the large-type portions read by the students. The tutor works with the student to sound out the phonemically regular words, asks comprehension questions about the whole story, and has the student reread passages out loud to gain fluency. Writing activities are also incorporated into the reading activities.

As noted, the tutoring model is closely integrated with the reading program (Slavin, Madden, Karweit, Livermon, & Dolan, 1990), in which students are regrouped according to their reading levels. Use of tutors as reading teachers allows schools to reduce class size to about 15 students who are all at one level, so there are never multiple reading groups in any reading class. This allows teachers to spend the entire class period actively teaching reading, as it removes the need for the follow-up or seatwork activities typical of classes with multiple reading groups. The beginning reading program emphasizes reading to students, engaging students in discussions of story structure, and developing oral language skills. Students begin using shared stories, as described earlier. As letter sounds and sound blending strategies are taught, students can apply them in their books. Students do a great deal of partner reading and pair practice activities, and writing is taught along with reading.

There is a high degree of structure in the beginning reading program, which is helpful in integrating the classroom instruction with the tutoring session. Expectations for each lesson are clear, so the teacher and tutor can know that they are working on the same objectives. Integration is also facilitated by the use of brief tutor/teacher communication forms, on which each can tell the other about particular successes or problems a child is experiencing.

Success for All is currently being evaluated in several schools in several school districts in six states. Evaluations most relevant to the tutoring aspect of the program relate to low achievers in two Baltimore schools that have had adequate funding to provide a high level of tutoring services for several years. Abbottston Elementary, the original pilot school, has implemented Success for All for 4 years. City Springs Elementary is a fully funded site whose implementation began a year after Abbottston. Each school was matched with a similar comparison school, and then students were individually matched on standardized reading measures. The student bodies at both Baltimore schools are almost entirely African American. Seventy-six percent of Abbottston's students qualify for free lunch. City Springs serves the most disadvantaged student body in the district; all its

READING RESEARCH QUARTERLY April/May/June 1993 28/2

Table 6 Effects of Success for All on low achieving students

	Abbottston site				City Springs site		
Measure	Year 1	Year 2	Year 3	Year 4	Year 1	Year 2	Year 3
Grade 1							
Woodcock Letter-Word	+0.42	+1.57	+1.09	+2.40	+0.08	+1.03	+0.57
Woodcock Word Attack	+1.34	+4.22	+1.00	+1.30	+0.51	+1.77	+0.71
Durrell Oral Reading	+0.99	+1.97	+0.21	+1.79	+1.14	+0.23	+0.37
Durrell Silent Reading	+1.30	+1.73	+1.06		+0.47	+0.45	
Mean	+1.01	+2.37	+0.84	+1.83	+0.55	+0.87	+0.55
Grade 2							
Woodcock Letter-Word		+0.39	+0.37	+1.07		+0.09	+0.98
Woodcock Word Attack		+0.66	+1.78	+1.28		+0.75	+1.36
Durrell Oral Reading		+0.52	+0.71	+0.87		+0.28	+0.98
Durrell Silent Reading		+1.26	+0.64			+0.16	
Mean		+0.71	+0.88	+1.07		+0.32	+1.11
Grade 3							
Woodcock Letter-Word			+0.57	+1.22			+0.20
Woodcock Word Attack			+1.22	+2.70			+0.50
Durrell Oral Reading			+1.11	+1.82			+0.78
Durrell Silent Reading			+1.36				
Mean			+1.07	+1.91			+0.49

Note Data are effect sizes from Slavin, Madden, Karweit, Livermon, & Dolan, 1990; Slavin, Madden, Karweit, Dolan, & Wasik, 1990; April, Madden, Slavin, Karweit, Dolan, Wasik, Shaw, Leighton, & Mainzer, 1991; and Madden, Slavin, Karweit, Dolan, & Wasik, 1992.

children come from housing projects, and 96% receive free lunch. Both are Chapter 1 schoolwide projects. Each May, students are individually assessed on scales from the Woodcock Language Proficiency Battery (1984) and the Durrell Analysis of Reading Difficulty (1980).

The results for the students in Grades 1-3 who scored in the lowest 25% on the pretests are summarized in Table 6. The amount of tutoring received by these students varied depending on their needs: almost all received some tutoring, but in some cases they received 8 weeks, while some second or third graders may have received more than a year of daily tutoring.

The results shown in Table 6 indicate powerful effects of the combination of tutoring, curricular changes, and family support services used in Success for All. At both schools in all years, first-grade low achievers have scored better than their matched counterparts in control schools (mean effect size = +1.15). Second graders who started in Success for All in the first grade or earlier also scored substantially better than control students (mean effect size = +.82), as did third graders in the program for 3 years (ES = +1.16). These second- and third-year effects should not be compared with the second- and third-year effects of Reading Recovery; the Reading Recovery data relate to the *lasting* effect of a first-grade intervention, while those for Success for All relate to the effects of *continuing* intervention. Although effect sizes stayed at approximately the same level in second and third grades as in first, this is an indication of a growing effect. Because standard deviations increase each year, a constant effect size means a growing difference between experimental and control groups in grade equivalents or raw scores.

In addition to effects on reading achievement, all three schools substantially reduced assignments of students to special education for learning problems and essentially eliminated retentions (Slavin, Madden, Karweit, Livermon, & Dolan, 1992).

As with Reading Recovery, there are methodological limitations to research on Success for All that may affect the results. First, because only one school was involved in each comparison, school effects could account for part of the observed differences. Lack of random assignment of schools or students also could have affected the results.

The effects of Success for All were positive for the lowest achieving quarter of students involved as well as for the other students in the school (Slavin, Madden, Karweit, Dolan, & Wasik, 1990; Slavin, Madden, Karweit, Livermon, & Dolan, 1990; Madden et al., 1991, 1992). However, the effects for the higher achieving students must be ascribed to the curriculum and other program elements, as few of them received any tutoring. Also it is important to note that schools using Success for All without extra resources for tutoring also obtained very positive results, although not as positive as those for the fully funded schools (see Slavin et al., 1992). These schools used their existing Chapter 1 funds to provide some tutoring (almost all to first graders), but could not sustain the amount of tutoring provided to Abbottston and City Springs low achievers. A school in Philadelphia used a modified version of Success for All to work with limited-English-proficient (LEP) Cambodian students, and also found very positive outcomes for these students and for non-LEP students (Slavin & Yampolsky, 1991). The evaluation of Success for All shows the potential power of a tutoring program that is integrated with a structured reading program. Evaluations of additional years will be needed to determine whether the program's goal of success for every child is realistic. Follow-up studies are needed to determine the validity of the program's assumption that success through the elementary grades will have long-term consequences, but the data collected to date clearly demonstrate the program's effectiveness when used at the beginning of students' school careers.

Like Reading Recovery, Success for All articulates instruction with assessment. The Word Identification, Letter Identification, and Word Attack subtests of the Woodcock assess letter and word knowledge and phonics skills. The Durrell Oral Reading test asks children to read passages in a limited amount of time and answer comprehension questions. Given the emphasis on decoding, oral language, perceptual analysis, reading strategies, and prose comprehension, these measures correspond with the model of reading in the Success for All program.

Like children in Reading Recovery, children in Success for All may be more familiar with the items that they are being evaluated on than are the controls because of the relationship between instruction and assessment.

Finally, unlike the verification of Reading Recovery, the fidelity of program implementation in Success for All has not been documented. It would be important to know whether the model of reading is being appropriately implemented and if tutoring sessions look similar across sites. Also, if there is no consistency of delivery of instruction across tutoring sessions, what do the effect sizes mean? Qualitative implementation data need to be collected in order to validate the consistency in delivery of instruction and to determine how this translates into increased reading performance.

Prevention of Learning Disabilities

Prevention of Learning Disabilities is a program developed by the Learning Disorders Unit of the New York University Medical Center that identifies first and second graders who are at risk for school failure and provides intensive instruction before they begin to fall behind in basic skills. Students involved in the program are screened in first grade using an instrument (SEARCH) that focuses primarily on neurological indicators of learning disabilities and on perceptual and general immaturity. Using diagnostic information from SEARCH, first graders are given lessons either individually or in small groups that attempt to strengthen their areas of perceptual weakness. The instructional interventions, called TEACH, are designed primarily to build perceptual skills, such as recognition, discrimination, copying, and recall, and are administered by certified teachers in 30-minute sessions three to five times per week.

Model of reading

Unlike Reading Recovery and Success for All, the Prevention of Learning Disabilities tutoring program is based on a physiological view of learning and learning disorders. As to reading itself, Silver and Hagin's (1990) model is based on the assumption that reading is a "complex process that must be analyzed according to component skills in order to understand the learning difficulty." However, these authors take a very atomistic view of reading and teaching reading. In compartmentalizing these reading skills, the goals are to identify those components with which the child is having difficulty and to teach to those specific skills.

Silver and Hagin propose that children need to have four skills in order to read: prereading skills, word attack skills, comprehension, and study skills. Prereading skills include the visual discrimination of letters, recognition of symbols in their correct orientation, the ability to organize symbols into groups, and several auditory skills. Word attack skills involve not only the use of phonics to figure out words, but also the identification of whole words using visual cues such as letter combinations. Comprehension involves having a rich vocabulary, being able to select the right meaning of a word, and making inferences. Study skills are described as the tools for acquiring information. These skills enable children to locate and select relevant elements within a sequence and organize the content of the text for later recall based on the goal of reading.

Although Silver and Hagin proposed these components of reading, not all of these aspects are directly taught in the program. There is considerable emphasis on matching, copying, and recalling individual letters and words, and little emphasis on reading for meaning. Phonics are not systematically presented, but instead letter-sounds are reinforced. In total, the Prevention of Learning Disabilities program includes in its model the following components of reading: perceptual analysis of print, decoding, and oral language proficiency.

Table 7 Effects of Prevention of Learning Disabilities on at-risk students

Measures	Effect sizes			
Silver & Hagin, 1979; Silver et al., 1981		End of Grade 1	End of Grade 2	End of Grade 3 Follow-up
SEARCH (Perception)		+.99	—	—
WRAT (Oral Reading)		+.85	+1.06	+.95
Woodcock Work Identification		+.94	+.91	+1.38
Woodcock Word Attack		+1.39*	+1.67	+1.26
SRA Comprehension		—	+.95	—
Gates-MacGinitie Comprehension		—	—	+.30
Gates-MacGinitie Vocabulary		—	—	+.15
Arnold et al., 1977	**TEACH vs. control**		**Reg. tutoring vs. control**	
WRAT	End of Grade 1	End of Grade 2	End of Grade 1	End of Grade 2
	+.33	+1.09	+.16	+.11
Mantzicopoulos et al., 1990	**TEACH vs. control**		**Phonetic vs. control**	
Total reading achievement (combined SAT, CTBS, CAT)	End of Grade 1	End of Grade 2	End of Grade 2	End of Grade 2
	+.16	+.21	+.28	+.13

Structure of tutoring

No coordination with the regular reading program is apparent in program descriptions. Children come to tutoring for 30 minutes, 3 to 5 days a week. Tutors, who are certified teachers, work on perceptual skills such as recognition, discrimination, copying, and recalling of information. There is no emphasis on reading connected text and no systematic presentation of phonics.

Results

An evaluation of Prevention of Learning Disabilities was conducted in inner-city New York City classrooms (Hagin, Silver, Beecher, 1978; Silver & Hagin, 1979). Students were randomly assigned to experimental or control classes, and those in the experimental group received TEACH instruction for 2 years. Table 7 summarizes the findings. On reading measures as well as on perception measures, the experimental students performed substantially better than controls. In the same study, Silver and Hagin (1979) found that students who had a full year of TEACH performed better than those who had only a half year.

In a similar study, Silver, Hagin, and Beecher (1981) found that third graders who received the TEACH intervention in first and second grade showed significantly greater performance in oral reading, word identification, and word attack skills (a measure that assesses the ability to sound out words) when compared to a no-treatment control group.

Arnold et al. (1977) replicated the Prevention of Learning Disabilities program in inner-city and middle-class schools in Columbus, Ohio. Using SEARCH, 86 first graders were identified as being at-risk for reading problems and were assigned to one of three groups: the TEACH intervention group, a group who received academic tutoring from a teacher, and a no-treatment control group. Students in the TEACH and regular tutoring group received tutoring for 30-minute sessions twice a week. Table 7 summarizes the findings. At the end of one year, the effects for both the TEACH intervention and the regular tutoring were minimal on the WRAT achievement test. However, at the end of the second year of the intervention, students in the TEACH group performed significantly better than the students in the regular tutoring and the no-treatment control group on the WRAT.

A more recent study by Mantzicopoulos, Morrison, Stone, and Setrakian (1990) found few effects for the TEACH intervention. In this study, first graders who were identified as at risk for reading failure by the SEARCH screen were assigned to three groups: a TEACH group, a phonics tutoring group, and a no-contact control group. In the phonics tutoring group, students were given phonics instruction, were drilled in phonics, and read phonemically regular books. This is in contrast to the TEACH group, which worked on visual-auditory discrimination activities. In both the TEACH and phonics tutoring groups, students received one-to-one tutoring for 30-minute sessions twice a week. The findings are summarized in Table 7.

On reading measures and perceptual measures, students in the TEACH group did not perform any differently than the phonics tutoring group or the no-contact controls. Not surprisingly, the phonics tutoring group did show some significant improvement in word attack skills, compared to the no-contact control.

Mantzicopoulos et al. (1990) suggest that one reason for the disappointing effects of TEACH was that the high attrition rate of their students produced a skewed sample distribution. Attrition, of course, is a factor in working with at-risk populations.

As in Reading Recovery and Success for All, the measures used to assess this program are consistent with the model of reading and the instruction delivered in tutoring. The Word Identification and Word Attack subtests of the Woodcock and the Gates-MacGinitie Vocabulary test assess letter, letter-sound knowledge, and word knowledge. The WRAT (oral reading), SRA comprehension, and the Gates-MacGinitie Comprehension scales assess reading connected text and comprehension, although these are multiple-choice tests and do not assess on line reading as does the Text Reading Level of Reading Recovery and the Durrell Informal Reading Inventory. Students in Prevention of Learning Disabilities are provided with minimal instruction in reading and answering comprehension.

A final concern about the findings from Prevention of Learning Disabilities is that, as with Success for All, there is no information about program implementation. There is no way of determining if the program implemented by Mantzicopoulos et al. (1990) was the same as Silver and Hagin's program or if variability in program implementation produced these different results. Qualitative implementation data need to be collected in order to determine if there is consistency across tutoring sessions and if what is proposed in the model of reading is being carried out in instruction.

Wallach Tutoring Program

The Wallach Tutorial Program (Wallach & Wallach, 1976) is, like Reading Recovery and Success for All, based on the idea that students who fail to learn to read in first grade are seriously at risk, and that carefully structured tutoring intervention can prevent reading failure. In this model, students receive 30 minutes of tutoring per day for a year. Unlike Reading Recovery and Success for All, the Wallach model uses paraprofessionals as tutors. The tutoring is directed to students who score below the 40th percentile on a standardized reading test.

Model of reading

According to Wallach and Wallach (1976), reading is a skill that "can be broken down into component parts; that these parts can be arranged in a cumulative, hierarchical manner such that learning of the latter parts builds systematically upon what has been learned already" (p. 56) and can be best learned by "systematically cumulating the mastery of component subskills" (p. 77). Wallach and Wallach propose that in acquiring these subskills, children must first establish competence "in the recognition and manipulation of sounds," and then acquire skill "in the use of the alphabetic code and in blending." Finally, they need to effectively apply "these competencies in reading printed material."

Because of this skills-mastery approach to reading, phonics are systematically presented in this program. Unlike Success for All, the phonics lessons are not coordinated with emphasis on reading connected text. Instead, reading comes after the letter sounds have been learned. Also, little consideration is given to metacognitive strategies. From Wallach and Wallach's point of view, the goal is to teach the students skills that they need to be readers. No attention is given to finding out the kinds of strategies the students are using and teaching new, more successful strategies.

Table 8 Grade equivalent dfferences and effect sizes for Wallach Tutoring Program

Measures	Grade equivalent differences	Effect sizes
	Tutored vs. matched control	
Wallach & Wallach (1976)		
Spache Word Recognition	*+.5	+.64
Spache Consonant Sounds Test	—	+.66
	Tutored vs. control group	
Dorval, Wallach, & Wallach (1978)		
Spache Word Recognition	*+1.6 to 1.8	—
Spache Reading Passages	—	—
CTBS	—	+.75

*Computation based on median scores

Much of this emphasis on the relationship of sound-symbol is a response to a finding in an earlier study (Wallach, Wallach, Dozier, & Kaplan, 1977), which indicated that at the end of kindergarten, most of a sample of disadvantaged students but few middle-class students had difficulty recognizing phonemes in words read to them, such as knowing that *man* but not *house* starts with the *mmm* sound. Wallach and Wallach argue that disadvantaged students need to be explicitly taught letter sounds so they can serve as a foundation for acquiring other skills necessary for learning to read.

In total, Wallach and Wallach includes only the following components of reading in their model: perceptual analysis, decoding, and oral language proficiency. They apparently believe that once the code has been cracked, oral language processes take over.

Structure of tutoring

The Wallach and Wallach program has three parts. For about 10 weeks, children are taught to recognize initial phonemes in words read to them, to recognize letters, and to associate letters and phonemes. In the second stage, students spend 2 to 3 weeks learning to sound out and blend easy words. For the remainder of the year, the children learn to apply their skills to classroom reading materials. Thus, the Wallach model begins as a completely separate tutoring program (like Reading Recovery) but later begins to integrate tutoring with classroom instruction (like Success for All).

Two studies have evaluated the Wallach model. The results of these studies are summarized in Table 8.

Table 9 Effects of Programmed Tutorial Reading

	Effect sizes			
	Programmed tutoring		Directed tutoring	
Measures	15 min. per day	30 min. per day	15 min. per day	30 min. per day
Ellson et al. (1968)				
Ginn Total Vocabulary	+.09	+.57	+.23	-.07
Ginn Total Comprehension	+.13	+.53	+.10	-.21
Ginn Total Word Analysis	-.19	+.46	+.28	-.01
Stanford Achievement Test	+.01	+.18	+.41	-.17
Ellson et al. (1965)				
Total Ginn Score	+.33			
Total Word Analysis Score	+.36			
Word Recall Score	+.78			
McCleary (1971):				
Ginn Achievement (all students)	+.40			
Ginn Achievement (low achievers only)	+.37			

The first evaluation was a field test in two inner-city Chicago schools (Wallach & Wallach, 1976). First graders who were identified at the beginning of the school year as low in "academic readiness" were randomly assigned to either tutoring or a no-treatment control.

On the Spache Word Recognition Scale, the tutored students scored 5 months higher than the control (G.E. 1.8 vs. 1.3) with an effect size of +.64. On the Spache Consonant Sounds Test, the tutored students also outperformed the control group, with an effect size of +.66. On the Spache Reading Passage scales, there were apparent differences favoring the tutored students but these were obscured by a floor effect on the test (which does not measure below a grade equivalent of 1.6).

A second study (Dorval, Wallach, & Wallach, 1978) evaluated the program in rural Roanoke Rapids, North Carolina. Students who received the tutoring were compared to similar students in the same school the previous year, to similar students in a comparison school who received the services of a full-time reading aide in their regular reading class, and to other students in the same comparison school who received neither tutoring nor aides. At the end of the year, students took the group-administered CTBS and were individually assessed on the Spache Word Recognition and Reading Passages scales. The various control groups did not differ from one another, so they can be pooled.

On the Spache Word Recognition Scale, the tutored students scored 8 months higher than controls (grade equivalent 2.3 vs. 1.5). On the Spache Reading Passages, the tutored students were reading at a median grade equivalent of 1.8, while control students were at a median of 1.6., but again a floor effect may account for this small difference. On the CTBS, tutored students scored at the 56th percentile, comparison students at the 34th, for an effect size of +.75.

Like the other programs, the measures used in assessing the Wallach and Wallach program match the model of reading and instruction. The Spache Word Recognition Test and the Spache Consonant Sound Test correspond to the emphasis on perceptual analysis and decoding. The CTBS also has sections that require word identification, analysis, and comprehension.

Again, there are no implementation data on the Wallach tutoring program. It would be important to know if the different effect sizes found in Wallach and Wallach (1976) compared to Dorval, Wallach, and Wallach (1978) are the result of implementation differences or other program factors. The differences found on the Spache Word Recognition test and Spache Reading Passages in the two studies is large enough that differences in program implementation need to be considered.

Programmed Tutorial Reading

Programmed Tutorial Reading is a highly structured tutoring program used with first graders who are in the lowest quartile on standardized reading tests. The program was originally developed by Douglas Ellson at Indiana University. The tutors for the program are paid paraprofessionals, volunteers, or parents. Students are tutored 15 minutes per day as a supplement to regular classroom instruction.

Model of reading

Ellson and his colleagues describe reading as " a complex activity that, at minimum, includes oral or sight-reading, phonics, and comprehension" (Ellson, Barber, Engle, & Kampwerth, 1965, p. 80). They describe a hierarchy in the important components of teaching reading in which sight reading has priority, followed by a phonemic analysis and synthesis, which they describe as not being necessary at the start, and, finally, comprehension of the meaning of visually presented words or sentences.

Of all the programs described, Programmed Tutorial Reading is the least explicit regarding its model of reading. This is partly due to the fact that Ellson and his colleagues were primarily interested in testing the structure of this individualized programmed instruction and intended to extend this model to other content areas such as math. The only components of reading that appear to be part of this model are perceptual analysis and decoding. They present no clear explanation or indication of how comprehension is taught.

The curriculum in Programmed Tutorial Reading is designed on the principles of operant conditioning. Students proceed by mastering small, sequential steps in the reading process and are reinforced for correct responses. The primary emphasis is on acquiring sight words. Phonics is also systematically presented in the context of acquiring words. There is no emphasis on reading words in connected text with the goal of learning new words and learning to comprehend what is read. Mastering individual components is expected to build a repertoire of behaviors that coordinate into reading. How this occurs is not explicitly stated.

Structure of tutoring

Students are cycled through a sequence of lessons on sight-reading, comprehension, and word analysis, which is repeated many times. Tutors are trained in specific strategies to present items, reinforce students for correct responses, and route students through the materials according to their responses. There is no coordination between tutoring and classroom instruction.

Results

Several studies have evaluated Programmed Tutorial Reading, but only three of these have compared the program to control groups over meaningful time periods with nonretarded populations. Table 9 summarizes the results of these studies.

Ellson, Harris, and Barber (1968) evaluated Programmed Tutorial Reading of two durations, over a full school year. Students were assigned to one of four tutored groups: Programmed Tutorial Reading for 15 minutes per day, Programmed Tutorial Reading for 30 minutes per day, an alternative tutoring program called Directed Tutoring for 15 minutes per day, and Directed Tutoring for 30 minutes per day. Then, a matched student was identified within the classroom of each tutored student. The students were first graders in 20 Indianapolis schools. Most of the schools served low-income populations, but the students were selected to be representative of their schools and did not necessarily have reading problems. The Directed Tutoring program did not use the programmed materials or highly structured procedures used in Programmed Tutorial Reading, but used remedial and supplementary materials more like those typically used in classrooms or in remedial reading programs.

The results (see Table 9) indicate strong effects of the 30-minute Programmed Tutorial Reading Program on tests provided along with students' Ginn basals (mean ES = +.52), but effects on the standardized Stanford Achievement Test were near zero, as were overall effects of the 15-minute per day program. Small positive effects were found for the 15-minute per day Directed Tutoring program, but (oddly) effects of the 30-minute Directed Tutoring treatment were slightly negative. Another study, by Ellson, Barber, Engle, & Kampwerth (1965), compared 15 minutes per day of Programmed Tutorial Reading for a semester to an untreated control group. In this case, moderate positive effects were found on the three measures used.

The largest methodologically adequate study of Programmed Tutorial Reading was done by McCleary (1971) in Lenoir County, North Carolina. In this study, low-achieving first graders were matched and assigned to experimental or control groups. The experimental students were tutored for the entire school year for 15 minutes per day. Positive effects on the Ginn reading test were found for the sample as a whole (ES = +.40) and for the poorest readers (ES = +.37). In addition, retentions in first grade were 55% lower in the tutored group than in the nontutored group. Taken together, the evaluations of Programmed Tutorial Reading suggest that the program has positive effects on student reading achievement, but the effects are smaller and less consistent than those for the programs that use certified teachers.

Two issues need to be addressed regarding the Program Tutorial Reading project. The study done by Ellson and his colleagues found different results than the McCleary study. These differences could be due to the fact that McCleary had a better experimental design than Ellson et al. However, another explanation which should be considered is that the differences found were the result of differences in implementation of the program at different sites. This is only speculative since implementation data are not available on either study.

In Programmed Tutorial Reading, the assessment used to measure outcomes matches the model of reading in the program. The Ginn Total Word Analysis and Vocabulary tests assess word identification and decoding skills, two skills that are emphasized in this program. The Ginn Total Comprehension consists primarily of reading short sentences or passages and answering questions. This section of the Ginn primarily assesses lower level comprehension skills that are taught specifically in this program. Unlike the comparison group, students in the Programmed Tutorial Reading group were taught tasks in tutoring which were similar to those used in evaluating the program.

Discussion

One-to-one tutoring of low-achieving primary-grade students shows potential as an effective instructional innovation. Across 16 separate studies of cohorts involving five different tutoring methods, effect sizes were substantially positive in nearly every case.

The five tutoring programs discussed here vary enormously in models of reading, curriculum, tutoring methods, duration, integration with regular classroom instruction, and other characteristics. The studies are equally diverse in populations, measures, and procedures. However, some patterns can be perceived.

First, programs with the most comprehensive models of reading, and therefore the most complete instructional interventions, appear to have larger impacts than programs that address only a few components of the reading process. Reading Recovery and Success for All include in reading instruction several components of reading such as perceptual analysis, conventions of print, error correction strategies, decoding, comprehension, error detection, and reading strategies. Moreover, they have comprehensive approaches to teaching the complex process of reading. In contrast, the Prevention of Learning Disabilities program which focuses only on building specific skills related to the reading process produced less consistent comprehension outcomes.

Second, it is not enough that programs simply use tutors. The content of the reading program in addition to the form of instructional delivery may be important variables. Ellson et al. (1968), for example, found the Programmed Tutorial Reading model to be significantly more effective than a standard "directed tutoring" intervention, and Arnold et al. (1977) found the Prevention of Learning Disabilities (TEACH) program to be considerably more effective than "regular tutoring." Mantzicopoulos et al. (1990) failed to replicate the findings of the earlier studies of Prevention of Learning Disabilities, but similarly found few effects of a "standard" phonics-based tutoring approach. An Ohio statewide study of Reading Recovery failed to find any positive effects of two alternative models of one-to-one tutoring (Pinnell et al., 1991). These findings, plus the apparent advantage of tutoring by certified teachers over tutoring by paraprofessionals, provides support for the proposition that for tutoring to be maximally effective it must improve the quality of instruction, not only increase the amount of time, incentive value, and appropriateness to students' needs (see Wasik & Slavin, 1990).

Third, programs using certified teachers as tutors appeared to obtain substantially larger impacts than those using paraprofessionals. Effect sizes for Programmed Tutorial Reading and the Wallach Tutorial Program generally fell in the range of +.20 to +.75, while those for the programs using certified teachers produced average effects from +.55 to +2.37 by the end of first grade. The teacher-delivered and paraprofessional-delivered models also differed in curriculum. Both the Wallach model and Programmed Tutorial Reading use highly structured, clearly described instructional materials, which in the latter program were explicitly patterned on programmed instructional methods usually designed for self-instruction. In contrast, the three teacher-administered models rely on teachers' judgment, flexibility, and knowledge of how children learn.

Only one program, Success for All, is designed to integrate completely with regular classroom instruction, and this program also produced some of the largest effect sizes. Although coordinating the tutoring sessions with classroom instruction is sensible in theory, empirical data need to be collected to determine its importance. The type of classroom instruction with which the tutoring was coordinated would also be an important factor. In addition, lack of consistency between how reading is presented in the classroom and how it is presented in tutoring may present a mismatch in the way reading is taught and result in confusion for the children. However, if Reading Recovery and Programmed Tutorial Reading were used both in the classroom and in tutoring, Reading Recovery might still have greater effects because its model of reading and delivery of instruction may be more effective. All of this remains to be determined in additional studies.

Several studies evaluated the *cumulative* and *lasting* effects of one-to-one tutoring in the early grades. Studies of two Success for All schools (Slavin et al., 1992) found that as students continued into second and third grades, initial positive effects continued to grow. Similar cumulative effects were found for Prevention of Learning Disabilities in two studies (Silver & Hagin, 1979; Arnold et. al., 1977) but not in a third (Mantzicopoulos et al., 1990). Silver & Hagin (1979) also found that students who experienced Prevention of Learning Disabilities for a full year learned more than those who had it for a semester, and Ellson et al. (1968) found that gains were greater when students received 30 minutes per day of Programmed Tutorial Reading than when they received only 15 minutes.

Because one-to-one tutoring (especially by a certified teacher) is expensive, the lasting effects of this approach are of great importance. Reading Recovery has been evaluated for lasting effects, and the results are positive but complex. On one hand, the raw score gains that students made on Text Reading Level in first grade have maintained through the end of third grade in two different cohorts (DeFord et al., 1988; Pinnell, 1988). On the other hand, because standard deviations of this measure increase each year, effect size estimates have diminished each year for both cohorts. A 1-year follow-up of Prevention of Learning Disabilities showed consistently positive effects for the third graders for most measures with the exception of performance on the Gates-MacGinitie Comprehension test. The effects for reading comprehension decreased 1 year after the intervention.

Two of the tutorial programs, Success for All (Slavin et al., 1992) and Programmed Tutorial Reading (McCleary, 1971) documented substantial reductions in retentions as a result of first-grade tutoring, and Success for All (Slavin et al., 1992) also showed reductions in special education referrals.

Is tutoring cost effective?

It should not come as a surprise that one-to-one tutoring of primary grade students is effective. A more important question is whether it is effective enough to justify its considerable cost. One way to address this question is to compare tutoring to other expensive interventions. For example, experiments in Tennessee, New York City, Toronto, and Indiana have reduced class size by almost half. This is the same as hiring an additional teacher for each class, who could instead be used to provide one-to-one tutoring for 20 minutes per day to about 15 students. The best and most successful of these class-size experiments, a Tennessee statewide study, found a cumulative effect of substantially reducing class size from kindergarten to third grade of about +.25 (Word et al., 1990), less than that found in any of the tutoring models. A follow-up study 1 year later found lasting effects of 4 years of small classes to be only +.13 (Word et al., 1990). Other studies of halving class size have found even smaller effects (Slavin, 1989). The effects of having aides work in the classroom have been found to be minimal in many studies (see Scheutz, 1980; Slavin, in press); the same aides could be used as tutors using models designed for that purpose, or could be replaced by teachers for a greater impact.

On the other hand, it is not yet established that a heavy investment in first grade will pay off in permanent gains for at-risk students. The Reading Recovery and Prevention of Learning Disabilities results hold out some hope for lasting gains, and the cumulative effects of Success for All also show promise for maintaining initial gains. Reductions in retentions and special education referrals, seen in two of the tutoring models, have both immediate and long-term impacts on the costs of education for low achievers. Substantial savings due to reduced retentions and special education placements have been shown for Reading Recovery (Dyer, 1992) and for Success for All (Slavin et al., 1992). However, if first-grade tutoring models prove to have long-term effects either without additional intervention (as in Reading Recovery) or with low-cost continuing intervention (as in Success for All), cost effectiveness will not be the only criterion for deciding to use these models. For if we know that large numbers of students can be successful in reading the first time they are taught, and that the success not only lasts but also builds a basis for later success, then educators and legislators may perceive an obligation to do whatever it takes to see that all students do in fact receive that which is necessary for them to succeed.

Future research

In many ways, research on preventive tutoring models is in its infancy. Although the studies reviewed here clearly indicate a strong positive effect of well-designed tutoring models, there are many important issues to be understood.

On the programmatic side, one important set of questions concerns how much reading failure can be prevented using resources short of one-to-one instruction by certified teachers. Could one-to-two or one-to-three instruction be nearly as effective? Could forms of tutoring using paraprofessionals be devised that would be nearly as effective as forms requiring certified teachers? Must tutoring be done daily, or could it be done less frequently? How much time must be allotted to tutoring each day? These issues need to be empirically tested.

More work is clearly needed on long-term effects of tutoring, not only on achievement but also on special education referrals and need for long-term remediation, critical elements in any consideration of cost effectiveness. Also, studies of alternative approaches to tutoring are needed. Successful models range from the phonemic, rigidly prescribed Programmed Tutorial Reading, to the "learning to read by reading" emphases of Reading Recovery and Success for All, to the focus on specific perceptual deficits of Prevention of Learning Disabilities. In the studies, it may be that each of these types of approaches would be successful with different children, and that someday we may know which type of program will work best with children of a given profile.

The issue of selection of assessment measures based on what is being taught in the programs has been discussed. Recently, researchers have been calling for more authentic measures for assessing what children learn. One possible way of trying to establish some understanding across programs would be to assess children in each program on the same measures, both standardized tests and perhaps more importantly, ongoing literacy performance measures (see Taylor, 1990). This information would allow some cross-program comparisons and also help in determining generalizability of what is taught to other forms of assessment.

A great deal of work is needed to understand why tutoring is effective. The rudimentary explanation offered in this article must be replaced by a far more sophisticated understanding of cognitive and motivational processes activated in tutoring that are not activated to the same degree (at least for at-risk children) in the regular classroom. Understanding how at-risk children learn to read in tutoring would contribute to an understanding of how at-risk children learn in general; the tutoring setting provides an ideal laboratory in which the process of learning to read can be observed as it unfolds over time. Microanalysis of tutor/child discourse could contribute to our understanding of how children learn to read (Green & Weade, 1985; Handerhan, 1990).

This qualitative understanding of tutoring would also help address the important issue of implementation across tutoring sessions. Only Reading Recovery has attempted to assess implementation and the effect this has on outcome data. Understanding how instruction is delivered will also help in tutor training. Every tutor is different and brings to the tutoring session his or her own unique understanding of that child and reading. However, each program has specific prescribed theories of reading and how these theories translate into practice. The goal is to ensure that instruction is in concert with the model of reading and is consistent from one tutoring session to the next.

As discussed in the introduction of this article, it would have been helpful to discuss the differences in each program's theory of reading. However, this was not possible because only Reading Recovery has made attempts at outlining a clear, coherent theory of reading. Instead, these programs take a pragmatic approach; that is, the evaluations focus on producing data to indicate that the programs work, not *why* they work. However, articulating a theory of reading based on empirical evidence is a valuable contribution to the field of reading. This area is ripe for theory development. It would be important to begin to understand how the interaction between the tutor and the student results in learning to read. Clarifying a theory of reading would add to a fundamental understanding of why the components included in a particular program make the program effective.

Finally, several limitations to this type of research synthesis need to be addressed. First, when only tutoring programs are reviewed, research on other effective interventions for preventing reading failure is not addressed. Other classwide reading programs also have shown some effectiveness. Also, in a best evidence synthesis, programs are grouped together and examined in terms of effectiveness. However, each program has very distinct characteristics and has been tested on different populations. Although we have attempted to look at some specific similarities and difference among programs, a review of this kind does not examine the nuances of each program nor does it address the qualitative differences that exist in the tutor-child dyad. Also, to test the relative effectiveness of these programs, studies need to be conducted in which children are randomly assigned to alternative programs and a control group, and in which the children's success on a variety of measures is assessed.

Although we want to know much more about how tutoring works and how to maximize its effectiveness (and minimize its cost), it appears from the research reviewed in this article that one-to-one tutoring is a potentially effective means of preventing student reading failure. As such, preventive tutoring deserves an important place in discussions of reform in compensatory, remedial, and special education. If we know how to ensure that students will learn to read in the early grades, we have an ethical and perhaps legal responsibility to see that they do so. Preventive tutoring can be an alternative for providing a reliable means of abolishing illiteracy among young children who are at risk for school failure.

REFERENCES

ARNOLD, L.E., BARNEBEY, N., MCMANUS, J., SMELTZER, D.J., CONRAD, A., & DESGRANGES, L. (1977). Prevention of specific perceptual remediation for vulnerable first-graders. *Archives of General Psychiatry, 34*, 1279-1294.

BAUSELL, R., MOODY, W., & WALZL, F. (1972). A factorial study of tutoring versus classroom instruction. *American Educational Research Journal, 9*, 591-597.

BLOOM, B.S. (1981). *All our children learning*. New York: McGraw-Hill.

CARROLL, J.B. (1963). A model for school learning. *Teachers College Record, 64*, 723-733.

CARTER, L.F. (1984). The sustaining effects study of compensatory and elementary education. *Educational Researcher, 13*, 4-13.

CLAY, M.M. (1979). *Reading: The patterning of complex behavior*. Portsmouth, NH: Heinemann.

CLAY, M.M. (1985). *The early detection of reading difficulties*. Exeter, NH: Heinemann.

CLAY, M.M., & CAZDEN, C. B. (1990). A Vygotskian interpretation of Reading Recovery. In L. Moll (Ed.), *Vygotsky and education*. New York: Cambridge University Press.

CLOWARD, R.D. (1967). Studies in tutoring. *Journal of Experimental Education, 36*, 14-25.

COUNCIL OF CHIEF STATE SCHOOL OFFICERS. (1987). *Elements of a model state statute to provide educational entitlements for at-risk students*. Washington, DC: Author.

COUNCIL OF CHIEF STATE SCHOOL OFFICERS. (1989). *Success for All in a new century*. Washington, DC: Author.

DEFORD, D., PINNELL, G.S., LYONS, C., & YOUNG, P. (1988). *Reading Recovery: Volume IX, Report of the follow-up studies*. Columbus, OH: Ohio State University.

DEVIN-SHEEHAN, L., FELDMAN, R. S., & ALLEN, V.L. (1976). Research on children tutoring children: A critical review. *Review Education Research, 46*, 355-385.

DORVAL, B., WALLACH, L., & WALLACH, M.A. (1978). Field evaluation of a tutorial reading program emphasizing phoneme identification skills. *The Reading Teacher, 31*, 784-790.

Preventing early reading failure

DYER, P. (1992). Reading Recovery: A cost-effectiveness and educational-outcomes analysis. *ERS Spectrum, 10*(1), 10-19.

EDMONDS, R.R. (1981). Making public schools effective. *Social Policy, 12*, 56-60.

ELLSON, D.G., BARBER, L., ENGLE, T.L., & KAMPWERTH, L. (1965). Programmed tutoring: A teaching aid and a research tool. *Reading Research Quarterly, 1*, 77-127.

ELLSON, D.G., HARRIS, P., & BARBER, L. (1968). A field test of programmed and directed tutoring. *Reading Research Quarterly, 3*, 307-367.

FARR, B., QUILLING, M., BESSEL, R., & JOHNSON, W. (1987). *Evaluation of PRIME TIME: 1986-1987. Final Report.* Indianapolis, IN: Advanced Technologies, Inc.

FRESKO, B., & EISENBERG, T. (1983). The effect of two years of tutoring on mathematics & reading achievement. *Journal of Experimental Education, 17*, 89-100.

GLASS, G., MCGAW, B., & SMITH, M.L. (1981). *Meta-analysis in social research.* Beverly Hills, CA: Sage.

GOODMAN, K., WATSON, B., & BURKE, C. (1987). *Reading Miscue Inventory*, 2nd Ed. New York: Richard C. Owen.

GREEN, J., & WEADE, R. (1985). Reading between the words: Social cues to lesson participation. *Theory into Practice, 24*, 14-21.

GREENFIELD, P.M. (1984). A theory of the teacher in the learning activities of everyday life. In B. Rogoff & J. Lave (Eds.), *Everyday cognition: Its development in social context.* Cambridge, MA: Harvard University Press.

GREENWOOD, C.R., DELQUARDI, J.C., & HALL, R.V. (1989). Longitudinal effects of classwide peer tutoring. *Journal of Educational Psychology, 81*, 371-383.

HAGIN, R.A., SILVER, A.A., & BEECHER, R. (1978). Scanning, diagnosis, and intervention in the prevention of learning disabilities: II. Teach: Learning tasks for the prevention of learning disabilities. *Journal of Learning Disabilities, 11*(7), 54-57.

HANDERHAN, E. (1990). Reading instruction as defined by "successful" teachers and their first-grade students within an early intervention programs. *Dissertation Abstracts International*, No. AAC910512.

HARRIS, P. (1974). *Programmed tutorial reading project.* (Submission to Joint Dissemination Review Panel). Washington, DC: U.S. Department of Education.

HUCK, C.S., & PINNELL, G.S. (1986). *The Reading Recovery Project in Columbus, Ohio. Pilot Year, 1984-85.* Columbus, OH: Ohio State University.

KENNEDY, M.M., BIRMAN, B.F., & DEMALINE, R.E. (1986). *The effectiveness of Chapter 1 services.* Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.

LLOYD, D.N. (1978). Prediction of school failure from third grade data. *Educational and Psychological Measurement, 38*, 1193-1200.

LYONS, C.A. (1991). A comparative study of the teaching effectiveness of teachers participating in a year-long or 2-week inservice program. In the 40th Yearbook of the National Reading Conference, *Learner factors/ teacher factors: Issues in literacy research and instruction.* Chicago: National Reading Conference.

LYONS, C., PINNELL, G.S., DEFORD, D., MCCARRIER, A., & SCHNUG, J. (1989). *The Reading Recovery Project in Columbus, Ohio. Year 3: 1988-1989.* Columbus, OH: Ohio State University.

MADDEN, N.A., & SLAVIN, R.E. (1989). Effective pullout programs for students at risk. In R.E. Slavin, N.L. Karweit, & N.A. Madden (Eds.), *Effective programs for students at risk.* Needham Heights, MA: Allyn & Bacon.

MADDEN, N.A., SLAVIN, R.E., KARWEIT, N.L., & LIVERMON, B. (1987). *Success for All: Teacher's manual for reading.* Baltimore, MD: Johns Hopkins University, Center for Research on Elementary and Middle Schools.

MADDEN, N.A., SLAVIN, R.E., KARWEIT, N.L., DOLAN, L.J., & WASIK, B.A. (1992, April). *Success for All: Longitudinal effects of a restructuring program for inner-city elementary schools.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

MADDEN, N.A., SLAVIN, R.E., KARWEIT, N.L., DOLAN, L., WASIK, B.A., SHAW, A., LEIGHTON, M., & MAINZER, K.L. (1991, April). *Success for All: Third year results.* Paper presented at the annual convention of the American Educational Research Association, Chicago.

MANTZICOPOULOS, P., MORRISON, D., STONE, E., & SETRAKIAN, W. (1990). *Academic effects of perceptually and phonetically based intervention for vulnerable readers.* Paper presented at the American Educational Research Association, Boston.

MCCLEARY, E. (1971). Report of results of Tutorial Reading Project. *The Reading Teacher, 24*, 556-559.

PINNELL, G.S. (1985). Helping teachers help children at risk: Insights from the Reading Recovery program. *Peabody Journal of Education, 62*(3), 70- 85.

PINNELL, G.S. (1988, April). *Sustained effects of a strategy-centered early intervention program in reading.* Paper presented at the annual convention of the American Educational Research Association, New Orleans.

PINNELL, G.S. (1989). Reading Recovery: Helping at-risk children learn to read. *Elementary School Journal, 90*, 161-182.

PINNELL, G.S., DEFORD, D.E., & LYONS, C.A. (1988). Reading Recovery: Early intervention for at-risk first graders. Arlington, VA: Educational Research Service.

PINNELL, G.S., LYONS, C.A., DEFORD, D.E., BRYK, A.S., & SELTZER, M. (1991). *Studying the effectiveness of early intervention approached for first grade children having difficulty in reading.* Columbus: Ohio State University, Martha L. King Language and Literacy Center.

PINNELL, G.S., SHORT, A.G., LYONS, C.A., & YOUNG, P. (1986). *The Reading Recovery Project in Columbus, Ohio. Year 1: 1985-1986.* Columbus, OH: Ohio State University.

POLLOWAY, E.A., CRONIN, M.E., & PATTON, J.R. (1986). The efficiency of group versus one-to-one instruction. *Remedial and Special Education, 7*(1), 22-30.

SCHEUTZ, P. (1980). *The instructional effectiveness of classroom aides.* Pittsburgh, PA: University of Pittsburgh, Learning Research and Development Center.

SCRUGGS, T.E., & RICHTER, L. (1985). Tutoring learning disabled students: A critical review. *Learning Disability Quarterly, 8*, 274-286.

SHAVER, J.P., & NUHN, D. (1971). The effectiveness of tutoring underachievers in reading and writing. *Journal of Educational Research, 65*, 107-112.

SILVER, A.A., & HAGIN, R.A. (1979). *Prevention of learning disabilities.* (submission to Joint Dissemination Review Panel). Washington, DC: U.S. Department of Education.

SILVER, A.A., & HAGIN, R.A. (1990). *Disorders of learning in childhood.* New York: John Wiley & Sons.

SILVER, A.A., HAGIN, R.A., & BEECHER, R. (1981). A program for secondary prevention of learning disabilities: Research in academic achievement and in emotional adjustment. *Journal of Preventive Psychiatry, 1*, 77-87.

SLAVIN, R.E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher, 15*(9), 5-11.

SLAVIN, R.E. (1989). Class size and student achievement: Small effects of small classes. *Educational Psychologist, 24*, 99-110.

SLAVIN, R.E. (in press). School and classroom organization in beginning reading: Class size, aides, and instructional grouping. In R.E. Slavin, N.L. Karweit, & B.A. Wasik (Eds.), *Preventing early school failure: Research, policy, and practice.* Boston: Allyn & Bacon.

SLAVIN, R.E., MADDEN, N.A., KARWEIT, N.L., DOLAN, L., & WASIK, B.A. (1990, April). *Success for All: Effects of variations in duration and resources of a schoolwide elementary restructuring program.* Paper presented at the annual convention of the American Educational Research Association, Boston.

SLAVIN, R.E., MADDEN, N.A., KARWEIT, N.L., DOLAN, L., & WASIK, B.A. (1992). *Success for All: A relentless approach to prevention and early intervention in elementary schools.* Arlington, VA: Educational Research Service.

SLAVIN, R.E., MADDEN, N.A., KARWEIT, N.L., LIVERMON, B.J., & DOLAN, L. (1990). *Success for All: First-year outcomes of a comprehensive plan for reforming urban education.* American Educational Research Journal, 27(2), 255-278.

SLAVIN, R.E., & YAMPOLSKY, R. (1991, April). *Success for All and the language minority student.* Paper presented at the annual convention of the American Educational Research Association, Chicago.

TAYLOR, D. (1990). Teaching without testing: Assessing the complexity of the children's literacy learning. *English Education, 22,* 4-74.

VON HARRISON, G., & GOTTFREDSON, C. (1986, Spring). Peer tutoring: A viable alternative to cross-age tutoring. *Contemporary Issues in Reading,* 125-131.

VYGOTSKY, L.S. (1978). *Mind in society: The development of higher psychological processes.* M. Cole, V. John-Steiner, S. Souber, & E. Souberman (Eds. and Trans.). Cambridge, MA: Harvard University Press.

WALLACH, L., WALLACH, M.A., DOZIER, M.G., & KAPLAN, N.E. (1977). Poor children learning to read do not have trouble with auditory discrimination but do have trouble with phoneme recognition. *Journal of Educational Psychology, 69,* 36-39.

WALLACH, M.A., & WALLACH, L. (1976). Teaching all children to read. Chicago: University of Chicago Press.

WASIK, B.A. & SLAVIN, R.E. (1990, April). *Preventing early reading failure with one-to-one tutoring: A best-evidence synthesis.* Paper presented at the annual convention of the American Educational Research Association, Boston.

WEADE, R., (1987). Curriculum 'n' Instruction: The construction of meaning. *Theory into Practice, 26*(1), 15-25.

WORD, E., JOHNSTON, J., BAIN, H.P., FULTON, B.D., ZAHARIAS, J.B., LINTZ, M.N., ACHILLES, C.M., FOLGER, J., & BREDA, C. (1990). *Student/Teacher Achievement Ratio (STAR): Tennessee's K-3 class size study, final report.* Nashville: Tennessee State Department of Education.

AUTHOR NOTES

An earlier version of this article was presented at the annual convention of the American Educational Research Association, Boston, Massachusetts, April, 1990.

This article was written under funding from the Office of Educational Research and Improvement, U.S. Department of Education (Grant No. OERI-R-117-R90002). However, any opinions expressed are those of the authors and do not represent OERI positions or policies. The authors thank Judith Green, Jim Byrnes, and the three anonymous reviewers for their helpful comments on earlier drafts of this article.

Responsive Practices in the Middle Grades: Teacher Teams, Advisory Groups, Remedial Instruction, and School Transition Programs

DOUGLAS J. MAC IVER and JOYCE L. EPSTEIN
Johns Hopkins University

In this article, we analyze data obtained from "Education in the Middle Grades," a national survey of practices and trends using a representative sample of principals in public schools that contain grade 7, to examine the use and perceived effects of practices that are believed by many educators to be especially responsive to the needs of early adolescents. These responsive practices include group advisory periods, interdisciplinary teacher teams, remedial instruction programs, and "school transition" activities. Multiple regression analyses suggest that grade organization is not a consistent determinant of responsive middle-grades practices. Overall, 7–9 junior high and 7–12 combination schools have fewer responsive practices than other middle-grade organizations. There are educationally significant but modest relationships between a school's use of responsive practices and principals' perceptions of the outcomes obtained by the school and its students. Different practices are associated with different indicators of school and student success. Principals report a stronger school program overall when they invest heavily in interdisciplinary teams of teachers to create supportive conditions for teachers and students. Principals expect fewer students to drop out before high school graduation when the school uses supportive advisory group activities or responsive remediation programs. Principals report that extensive school transition programs reduce the number of students who need to repeat the grade immediately following the transition. The implications of the results for the improvement of education in the middle grades are discussed.

> For many youth, early adolescence is one of the last real opportunities to affect their educational and personal trajectory. The middle grade school, one of the key socializing institutions for young adolescents, represents a critical "turning point" in the lives of American youth. [Jackson and Hornbeck 1989, p. 831]

© 1991 by The University of Chicago. All rights reserved.
0195-6744/91/9904-0010$01.00

August 1991

Middle School Grades

> The most successful [middle grade] schools . . . are those meeting the needs of early adolescents for security, support, and success in a proactive manner. [VAN HOOSE AND STRAHAN 1988, p. 26]

Early adolescents are characterized by a plethora of simultaneous and often conflicting needs (Epstein 1988; Van Hoose and Strahan 1988). For example, they need the security and support of close, caring adult supervision and guidance at the same time that they need increasing autonomy from adults. They need and want attention and recognition for their own unique abilities, successes, and achievements, but they also want to be part of a crowd. As they engage in the life-shaping process of self-exploration and self-definition, they need help in remedying their weaknesses and developing their strengths, but—in order to be effective—this help must be offered in a way that does not stigmatize them, label them, or separate them from their peers (Mac Iver and Epstein, in press).

Since the turn of the century, when G. Stanley Hall (1905) published the first major text on adolescence as a separate stage of development, adolescent psychology has been influencing the rationale, curricula, and pedagogy of middle-grades schools (Perlstein and Tobin 1988). For example, since the 1920s, some middle-grades schools have been implementing programs that they believe are especially responsive to early adolescents, including such practices as exploratory courses (Koos 1927; Smith 1925); homerooms and teacher advisories (Hieronimus 1917), and extracurricular activities (Briggs 1922; Kitson 1926). In the fifties, sixties, and seventies, key middle-level educators began advocating additional developmentally appropriate practices including (*a*) core curriculum approaches emphasizing the correlation of subject areas, the integration of learning across disciplinary boundaries, and interdisciplinary team teaching, (*b*) discovery and inquiry methods, (*c*) flexible scheduling, and (*d*) ungraded programs (Alexander and George, 1981; Mac Iver and Epstein, in press).

The middle-school movement—which now has almost three decades of experience behind it—is devoted to implementing responsive practices in the middle grades. This movement has met with mixed success.

DOUGLAS J. MAC IVER is associate research scientist and codirector of the Middle Grades Program at the Johns Hopkins University Center for Research on Effective Schooling for Disadvantaged Students (CDS). JOYCE L. EPSTEIN is principal research scientist and codirector of the Center on Families, Communities, Schools and Children's Learning and also codirector of the Middle Grades Program at CDS.

American Journal of Education

Although some middle-grades schools are far along the road to institutionalizing many of the practices listed above, others have not begun the restructuring process or have consciously rejected certain recommended practices (Epstein and Mac Iver 1990). One reason for the great diversity of educational practices and approaches currently found in middle-grades education is that there has been little useful research to help educators decide which practices are beneficial for early adolescent students and which are ineffective. The research that has been done has been limited in the location and nature of the samples of schools and students, the breadth and depth of information on middle-grades practices, or the comparisons of alternative organizations of middle-grades schools.

Out of the many responsive practices appearing on past or current lists of recommendations for education in the middle grades, this article focuses on (*a*) teacher advisory, homeroom, or group advisory programs; (*b*) interdisciplinary teams of teachers who share the same students and coordinate their instructional programs across subjects; (*c*) special remedial activities for students who fall behind or learn more slowly than other students; and (*d*) transition or articulation activities with students, parents, and school staff to ease students' transitions from one level of schooling to the next (i.e., from the elementary to the middle grades, and from the middle to the high school grades).

Few data exist concerning the prevalence or effects of such programs. This study examines the structure, use, and perceived effects of these four different types of responsive practices in a national sample of public schools that serve young adolescents. It considers differences in the use of these practices in schools with different grade spans, in different locations, and with different types of student populations. Further, the study uses principals' opinions, estimates, and best guesses to begin to address the question, Are the practices that are being implemented having positive effects on the strength of middle-grades programs and student outcomes? That is, it examines how these practices are related to principals' evaluations of their middle-grades program, to principals' predictions of the percent of current seventh graders who will not graduate from high school, and to other school-level outcomes such as retention rate.

Method

The 2,400 schools in the study are a probability sample of public schools in the United States having seventh-grade students. From the approximately 25,000 public schools that serve regular seventh-grade students, 2,000 schools were sampled with probabilities proportional to each school's enrollment per grade level. In addition, two subuniverses of schools were oversampled: schools serving both elementary and middle grades in metropolitan areas and schools in districts with substantial rates of poverty (i.e., Orshansky index at or above 25). Approximately 200 of each type were added to the sample, making the total sample size 2,400.

In the spring of 1988, the Johns Hopkins Center for Research on Elementary and Middle Schools (CREMS) sent survey forms by mail to the principals of the 2,400 schools in the sample. A total of 1,753 (73 percent) of the principals provided information on their school for this study, including 1,344 who returned surveys by mail and 409 who completed shorter telephone interviews. The telephone interviews were conducted with a random subsample of all nonrespondents to the mail survey. Weighting the telephone interview responses to account for the essentially similar nonresponding schools that were not followed up by telephone brings the weighted response rate to 93 percent for the items that were common to the mail and telephone surveys.

For data analysis purposes, each school was first assigned a "weight" that was the inverse of its probability of selection. This weighting returns the sample to an equal probability (representative) sample of schools. Then, because we wanted to describe the experience of the typical middle-grades student, each school was upweighted by the school's enrollment per grade level, scaled so that the weighted total number of schools is equal to the unweighted raw number of schools (1,344 for items asked by mail only and 1,753 for items asked over the phone and by mail).

Multiple regression analyses were used to identify significant antecedents and consequences of four sets of practices: group advisory periods, interdisciplinary teams, remedial instruction activities, and school transition programs. The variables used in these analyses are presented in the Appendix and are described in more detail later. The variables included measures of (*a*) practices, programs, policies, and staff in the middle grades (Appendix, variables I–XI); (*b*) characteristics of the school (variables XII–XIV); (*c*) characteristics of the school's students (variables XV–XVIII); and (*d*) outcomes obtained by the school and its students (variables XIX–XXIV). Throughout the questionnaire, principals were reminded to focus only on their school's practices and programs for middle-grades students ("students in grades 5 through 9") even if the school also contained students from other grade levels. The complete questionnaire is found in Epstein and Mac Iver (1990).

Grade-span categories.—Grade span was one school characteristic that principals reported (variable XII). The data indicate that public schools in the United States that enroll seventh-grade students include 29 different grade spans. Schools were categorized in six groups for analyses: (1) K–8 (schools that start with an elementary grade—usually kindergarten—and end with a middle grade—usually eighth); (2) K–12 (schools that start with an elementary grade and continue through twelfth grade); (3) 7–12 (schools that start with a middle grade and continue through twelfth grade); (4) middle schools (mainly 6–8 schools, but also 5–8, 5–7, and 6–7); (5) 7–8 schools; and (6) junior high schools (schools that start with a middle grade and continue through ninth grade). These categories were represented in the analyses by five dummy variables; middle schools served as the control category (the category coded zero on each dummy variable). The dummy variables were used to examine the degree to which the grade span of a school predicts implementation of group advisory periods, interdisciplinary team approaches, school transition programs, or innovative remedial instruction activities in the middle grades.

National patterns of grade organization.—More than half of the schools fall in the K–8 category or the middle school category (see table 1). Less common are 7–12, 7–8, K–12, and junior high schools. The number of schools in each category should not be confused with the number of students who attend the schools in that category. For example, although only about 45 percent of the school buildings that contain

TABLE 1

National Grade Span Patterns: Percentages of Public Schools of Different Types and the Percentage of Students Who Attend Each Type (*N* = 1,753 Schools)

Type of Grade Organization	Percent Schools	Percent Students
K–8 (and other elementary-middle combinations)	32.1	9.3
K–12 (and other elementary-middle-high combinations)	10.5	2.4
7–12 (and other middle-high combinations)	12.7	7.0
Middle school (mainly 6–8, but also 5–8, 5–7, and 6–7)	25.3	39.3
7–8	11.0	24.6
Junior high (mainly 7–9, but also 6–9 and 5–9)	8.4	17.4

grade 7 are middle, 7–8, or junior high schools, these schools are attended by over 80 percent of all seventh graders. In contrast, the most common type of school building, the K–8 school, enrolls only 9 percent of the nation's seventh graders.

Group Advisory Periods

In their attempt to offer early adolescents high-quality instruction from subject-matter experts, many schools establish departmentalized programs in which students receive instruction from a different teacher for each academic subject. However, when students change teachers every period, they may feel that there is no one teacher who really knows them, cares about them, or is available to help them with problems. To provide each student with a teacher who knows and cares about the student and is available as a mentor or advisor, many schools have established homeroom or group-advisory periods. About two-thirds of the schools in the United States that include grade 7 have one homeroom or group-advisory period, and 9 percent have two such periods (Epstein and Mac Iver 1990).

Although advisory or homeroom periods are common, many of the activities that occur during these periods are mechanical tasks (e.g., taking attendance, distributing notices, making announcements, orienting students to rules and programs) rather than social and academic support activities that use teachers' talents as advisors and that help students feel that someone is looking out for their interests and needs.

To explore the antecedents and consequences of using supportive activities during group-advisory periods, a composite was created indicating the mean frequency of occurrence of nine social or emotional support activities during a homeroom or group advisory period (see Appendix, variable I). These activities included meeting with individual students about problems, giving career information and guidance, discussing academic problems or issues, and similar activities. Principals indicated how frequently each activity occurred, using a five-point scale, ranging from 1 = never to 5 = daily. Schools not having a homeroom or group-advisory period were assigned a score of one on this variable to indicate that support activities never occurred during a group advisory period at these schools. The grand mean for this variable was 2.3 (SD = 1.2); each type of support activity occurred only a few times per year, on the average.

Antecedents and perceived consequences of using supportive activities during group advisory.—The first column in table 2 summarizes a multiple regression model in which the mean frequency of responsive activities

TABLE 2

Summary of Multiple Regression Analyses Exploring the Antecedents and Consequences of Using Supportive Activities during Homeroom/Advisory Period

Effect	Use of Supportive Activities	Strength of Guidance Programs	Percent of Boys Who Probably Will Not Graduate from High School	Percent of Girls Who Probably Will Not Graduate from High School
Grade organization:				
K–8 vs. middle school	.02	−.18**	−.08**	−.08**
K–12 vs. middle school	.02	−.10**	−.07**	−.06*
7–12 vs. middle school	−.05	.04	−.11**	−.09**
Junior high vs. middle school	−.12**	.01	−.04	−.02
7–8 vs. middle school	.00	−.02	−.01	.01
Region:				
West vs. Northeast	−.02	−.04	.12**	.12**
Midwest vs. Northeast	−.10**	−.02	.00	.01
South vs. Northeast	.02	−.02	.08*	.07
Percent black students	.08*	.03	.08**	.06
Percent families below poverty line	.08*	−.06	.13**	.14**
Percent professional families	.04	.08*	−.13**	−.11**
Average ability of students on entry	.00	.05	−.30**	−.30**
Population of SMSA	.07*	−.04	.07**	.10**
Use of supportive activities	. . .	.16**	−.06*	−.05*
Adjusted R^2	.05	.08	.31	.28

NOTE.—Table entries are standardized regression coefficients.
* $P < .05$.
** $P < .01$.

during advisory period is predicted based on (*a*) grade organization of school, (*b*) region, (*c*) the urbanicity of the area in which the school is located, (*d*) the percentage of black students in the school, (*e*) the percentage of the school's families whose income is below the poverty line, (*f*) the percentage of professional or managerial families in the school, and (*g*) the average ability of the students on entry to the school (Appendix, variables XII–XVIII).

Effects of grade organization.—The standardized regression coefficient of −.12 for junior high schools in the first column of table 2 indicates that these 7–9 schools use supportive group-advisory activities significantly less frequently than do 6–8 middle schools (the schools that served as the control category). Other grade organizations do not significantly differ from 6–8 middle schools in their use of supportive group-advisory activities. Other comparisons (not shown in table 2) indicate that junior high schools use supportive group-advisory activities significantly less often than every grade organization except for 7–12 schools. However, junior high and 7–12 schools are more likely than other schools to have at least one professional guidance counselor (Epstein and Mac Iver 1990) and thus may be less likely than others to perceive a need for a group-advisory program. Further, junior high and 7–12 schools are more likely than most other schools to have a large proportion of teachers who have secondary subject-matter certification (Epstein and Mac Iver 1990). Teachers who are secondary certified may feel poorly prepared to serve as teacher advisors. Most of their education focused on helping them become experts in their areas of specialization. Typically, they will have received less preparation than elementary-certified teachers in understanding and responding to students' nonacademic problems, interests, and concerns.

To test the hypothesis that the lower use of supportive group-advisory activities in junior high and 7–12 schools is due to the presence of professional guidance counselors and secondary-certified teachers, the regression analysis in the first column of table 2 was recalculated after adding "presence of guidance counselor" and "percentage of secondary-certified teachers" (Appendix, variables II and III) as predictors. As expected, schools with guidance counselors were less likely to use supportive group-advisory activities ($\beta = -.07$, $P = .03$). Similarly, the negative effect of having secondary-certified teachers on use of supportive group-advisory activities was nearly significant ($\beta = -.06$, $P = .06$). Nevertheless, the differences in use of supportive group-advisory activities between junior high or 7–12 schools and middle, 7–8, and K–8 schools did not lessen even after controlling for these two variables. Conversely, K–12 schools were no longer significantly different from junior high or 7–12 schools in use of supportive activities.

The finding that 7–9 junior high, 7–12, and now K–12 schools use supportive group-advisory activities less (even after taking into account these guidance counselor and certification effects) suggests that the inclusion of one or more of the high school grades in a school may make it less likely that the school will establish a strong group-advisory program for middle graders. Carnegie Unit requirements concerning course offerings (which begin in ninth grade) may limit the number and length of periods available for group-advisory activities in the high school years. Although there is nothing to prevent junior high or 7–12 schools from offering frequent group-advisory activities to their seventh and eighth graders (even if they cannot offer them to their ninth graders), many schools choose not to differentiate their program in this way.

Other antecedents.—There were regional differences in the use of supportive group-advisory activities. The −.10 coefficient for the Midwest in the first column of table 2 indicates that such activities occurred significantly less frequently in the Midwest than in the Northeast. In contrast, the West and the South did not significantly differ from the Northeast in use of these activities. Supportive group-advisory activities were used less in the Midwest than in any other region.

Finally, the frequency of supportive group-advisory activities increases as the percentage of black students in the school increases, as the percentage of families below the poverty line increases, and as the population of the schools' standard metropolitan statistical area (SMSA) increases. That is, schools with poor, predominantly black student populations in big cities are more likely than others to establish group-advisory periods that frequently provide social and emotional support to students.

Effect of supportive activities on perceived strength of guidance programs.—Next, we examined the possible consequences of providing supportive activities during homeroom or advisory periods. In schools where these activities often occur, are principals more likely to report that the school is meeting students' needs for guidance, advice, and counseling? Or do such activities make no discernible difference?

Principals rated the overall quality of their guidance and advisory program (Appendix, variable XIX) on a scale ranging from a high of 4 (signifying an excellent guidance program, in which present practices meet students' needs exactly) to a low of 1 (signifying a weak guidance program). The .16 coefficient in the second column (last row) of table 2 indicates that principals in schools with an advisory/homeroom program that features frequent use of supportive activities were significantly more likely than principals in other schools to rate their guidance program as strong.

August 1991

The grade-organization effects in the second column show that principals in K–8 and K–12 schools rated their overall guidance program as significantly weaker than did principals in middle schools. Finally, schools serving a high percentage of professional/managerial families rated their guidance programs as being stronger than did principals in other schools.

Effect of supportive activities on estimated dropout rates.—We asked principals to estimate the percentage of their current seventh graders who probably would not graduate from high school (Appendix, variable XX). One possible outcome of a strong homeroom/advisory program would be to reduce a schools' dropout rate below the rate one would otherwise predict based on the schools' grade organization, location, and type of student population. The final two analyses in columns 3 and 4 of table 2 indicate that principals in schools with more supportive homeroom/advisory activities do, indeed, report a significantly lower expected dropout rate for both boys and girls. These analyses also indicate that the expected dropout rate is higher in 6–8 middle schools than in K–8, K–12, and 7–12 schools, and higher in the West and South than in the Northeast. In addition, principals expect more students to drop out if their school is located in a big city, if their community contains many students living below the poverty line, and if their school serves many low-ability children or few children from professional families.

In sum, even after family and student background variables, region, and grade organization are statistically controlled, principals in schools with well-implemented group-advisory programs report that they have stronger overall guidance services and lower expected dropout rates. Although principals' estimates of the strength of their guidance services and of their future dropout rates are informative, they are imperfectly related to objective measures of guidance-program effectiveness and to actual dropout rates. Thus, it is important for future research to attempt to replicate these findings with more objective measures (e.g., once longitudinal data from the National Education Longitudinal Study sample [Haffner et al. 1990] are available, it will be possible for researchers to compare actual dropout rates for high school students who received or did not receive supportive group-advisory services when they were in the middle grades).

Interdisciplinary Teams

Many proponents of the middle-school philosophy view the establishment of interdisciplinary teams of teachers as the keystone of education in the middle grades (e.g., Merenbloom 1986; Vars 1987). They hypothesize that interdisciplinary teams will eliminate the isolation that many teachers feel by providing a working group of colleagues to conduct activities and discuss and solve mutual problems; that instruction will be more effective in schools that use interdisciplinary teaming because of increased integration and coordination across subjects; and that teachers on a team sharing the same group of students will be able to respond more quickly, personally, and consistently to the needs of individual students.

Our data indicate that about 42 percent of early adolescent students receive instruction from interdisciplinary teams of teachers sometime between grades 5 and 9. An interdisciplinary team most often consists of four teachers—a social studies teacher, an English teacher, a math teacher, and a science teacher—who share a group of 100–125 students (Epstein and Mac Iver 1990).

Implementation of an Interdisciplinary Team Approach

Schools vary in their level of implementation of interdisciplinary teacher teams. For example, in some schools, all students receive instruction from interdisciplinary teams of teachers, and team members are given a common planning period. In other schools, the team approach (if adopted at all) may be used with only a subset of the school's students (e.g., sixth-graders) and team members may not be given a common planning period. To measure the variation between schools in their emphasis on an interdisciplinary team approach to school organization and instruction, we created a composite variable (Appendix, variable IV) ranging from 0 (no use of interdisciplinary teams in the middle grades) to 3 (interdisciplinary teams and common planning periods at each of the middle grades in the school).

Implementation of a Departmental Approach

Some schools may choose to establish and emphasize departments instead of, or in addition to, interdisciplinary teams. These schools may organize their faculty by subject area, appoint department heads, give common planning periods to members of departments, and use disciplinary (single-subject) team teaching. A disciplinary organization and emphasis may be particularly welcomed by those teachers who prefer to identify with a department rather than an interdisciplinary team and who find it easier to collaborate with and learn from teachers who are in the same discipline. A school's commitment to departments was measured by a composite variable ranging from 0 to 3 (Appendix, variable V). A maximum score of 3 indicates that the school has departments organized with their own chairpersons or heads, has a common planning period for members of departments, and uses single-subject team teaching in each of the middle grades.

Antecedents of Interdisciplinary Teaming

The first column of table 3 summarizes a regression model that explores the antecedents of emphasizing an interdisciplinary team approach in the middle grades. The standardized coefficients indicating the effects of grade organization show that middle schools implement interdisciplinary teaming significantly more extensively than do other schools.

The coefficients associated with region (col. 1, rows 6–8) indicate that schools in the Northeast are more likely than schools in other regions to have adopted an interdisciplinary team emphasis.

It is interesting that schools that emphasize departments (with department heads, common planning periods for departments, and teacher teams within departments) are more likely than other schools to also organize and emphasize interdisciplinary teams (col. 1, row 14). This indicates that a departmental emphasis and an interdisciplinary team emphasis coexist in many schools.

Perceived Consequences of Implementing Interdisciplinary Team and Department Approaches in the Middle Grades

The second column in table 3 reports standardized regression coefficients from an equation predicting the strength of each school's overall middle-grades program (Appendix, variable XXI), based on its emphasis on interdisciplinary teaming, commitment to departments, and other variables. The significant positive coefficients in rows 14 and 15 suggest that a school's commitment to departments and its implementation of interdisciplinary teacher teams are both associated with increases in the strength of its overall program, according to principals' reports.

One other effect was significant. The higher the average ability of students in a school, the stronger the ratings given by the principal to the school's middle-grades program (row 12).

There is no evidence that the adoption of an interdisciplinary team approach or a commitment to departments reduces dropout rates. On the contrary, schools that emphasize interdisciplinary teaming have a

TABLE 3

Summary of Regression Analyses Exploring the Antecedents and Consequences of a School's Implementation of Interdisciplinary Teaming

Effect	Implementation of Interdisciplinary Teaming	Strength of Middle-Grades Program	Percent of Boys Who Will Probably Not Graduate from High School	Percent of Girls Who Will Probably Not Graduate from High School
Grade organization:				
(1) K–8 vs. middle school	−.07*	.01	−.08**	−.07**
(2) K–12 vs. middle school	−.07*	.05	−.07**	−.06*
(3) 7–12 vs. middle school	−.12**	−.02	−.09**	−.08**
(4) Junior high vs. middle school	−.26**	−.05	−.01	.00
(5) 7–8 vs. middle school	−.16**	−.05	.00	.02
Region:				
(6) West vs. Northeast	−.15**	.05	.10**	.10**
(7) Midwest vs. Northeast	−.12**	.03	.00	.02
(8) South vs. Northeast	−.15**	.08	.09*	.07
(9) Percent minority students	.02	.00	.12**	.11**
(10) Percent families below poverty line	−.04	−.01	.10**	.10**
(11) Percent professional families	.07	.03	−.13**	−.12**
(12) Average ability of students on entry	−.04	.19*	−.29**	−.29**
(13) Population of SMSA	−.02	.04	.05	.07**
(14) Commitment to departments	.08**	.16**	−.03	−.02
(15) Implementation of interdisciplinary teams	. . .	.13**	.08**	.07**
Adjusted R^2	.09	.06	.32	.29

NOTE.—Cell entries are standardized regression coefficients.
* $P < .05$.
** $P < .01$.

higher expected dropout rate than would be predicted based on background and demographic variables (cols. 3 and 4, row 15). Further, a school's level of commitment to departments is not a significant predictor of expected dropout rate (row 14).

Why is the relation between a school's emphasis on interdisciplinary teaming and the principal's reports of expected dropout rate positive? It may be that a school's dropout rate influences the school's openness to making a commitment to interdisciplinary teaming. Schools that have a historic pattern of high dropout rates may make stronger commitments to this and other promising practices in the hope of reducing these rates. At the time of the survey, principals in these schools may not yet have known whether using interdisciplinary teams of teachers would actually reduce the percentage of their students who would leave school before high school graduation.

An alternative explanation is that a focus on interdisciplinary teaming may divert schools from providing sufficient remedial and guidance services, which may be critical in dropout prevention. The data do not support this alternative hypothesis. Schools with a commitment to interdisciplinary teaming actually have more extensive remedial programs ($r = .12, P < .001$); provide more supportive group advisory activities ($r = .29, P < .001$); and have lower students-per-guidance-counselor ratios ($r = -.07, P < .05$). These correlations suggest that the original hypothesis may be correct. That is, schools with high dropout rates may often adopt interdisciplinary teaming and other responsive practices in their attempt to rescue potential dropouts.

Does Increased Common Planning Time and the Establishment of Team Leaders Help Teams Succeed?

One might assume that having sufficient common planning time to do collaborative work and having a team leader—someone who is directly responsible for coordinating and organizing team activities—would help an interdisciplinary team succeed. Yet, only 36 percent of the schools that use interdisciplinary teams give team members two or more hours of common planning time each week and less than 60 percent of all teams have a formal team leader (an elected or appointed leader, or a system in which team leadership rotates among members).

The correlations in table 4 suggest that the provision of adequate planning time and the establishment of team leaders make a real difference in principals' opinions of how a team functions and in what it accomplishes. For example, the amount of common planning time allocated to interdisciplinary teams (Appendix, variable VI) is strongly

TABLE 4

Zero-Order Correlations among Four Teaming-related Variables

	A	B	C	D
A. Amount of common planning time	. . .			
B. Have team leader? (0 = no, 1 = yes)	.31			
C. Proportion of common planning time spent on team activities: coordinate content, revise schedules, regroup students, etc.	.42	.39		
D. Benefits resulting from use of interdisciplinary teams.	.21	.18	.36	. . .

NOTE.—All correlations are statistically significant: $P \leq .001$.

associated with principals' estimates of the proportion of time that team members spend coordinating their activities (deciding common themes and related topics for instruction, altering schedules, regrouping students, discussing problems of specific students and arranging help, and so on [Appendix, variable VII, *b–g*]). Larger amounts of common planning time are also associated with obtaining greater benefits from interdisciplinary teaming according to principals' reports (Appendix, variable XXII). Similarly, as shown in column B of table 4, when interdisciplinary teams have formal leaders (Appendix, variable VIII), teams spend more of their common time engaged in team activities and produce greater benefits for their school.

Remedial Instruction Activities

All middle-grades schools have some students who fall behind or learn more slowly than others. The Carnegie Task Force on the Education of Young Adolescents (1989) recommends that all middle-level schools proactively address the needs of these students through remedial instruction activities that provide specialized instruction, extra coaching, and additional time to learn. We asked principals to report the remedial activities offered in their schools (Appendix, variable IX). Over 98 percent of the principals reported at least one program to help students who fell behind. The most common remedial activities were pull-out programs in reading or English (61 percent of the seventh graders attended a school offering such a program), after- or before-school coaching classes (58 percent), summer school (52 percent), and pull-out programs in math (51 percent). Schools were less likely to offer students an extra subject period in lieu of an elective or exploratory course (28 percent), and rarely offered remediation through Saturday classes (3 percent). Ironically, except for summer school, each of the special remedial activities listed was most common in schools where the average academic ability of students is considerably above the national norm (see table 5).

Antecedents of the Number of Remedial Programs Offered

The first column in table 6 reports standardized coefficients from a regression model that attempted to predict the number of remedial programs offered in each school. The adjusted R^2 of .03 for this model indicates that very little of the between-school variance in the extensiveness of remedial programs is explained by grade organization, region, and family and student background variables. Only three effects were significant. The number of remedial programs offered by a school is positively related to the average ability level of the school's students and the urbanicity of the area surrounding the school (rows 12 and 13). Also, middle schools offer significantly more remedial programs than do 7–12 schools (row 3).

Perceived Consequences of the Number of Remedial Programs Offered

Ideally, an extensive remedial instruction program should make it possible for a school to lower its retention rate (Appendix, variable XXIV). Our data suggest that, instead of serving as an alternative to retention, an extensive remedial program tends to go along with high rates of retention (see table 6, col. 2, row 16). Just as we saw with other indicators of responsive programs, schools with severe problems (e.g., a high number of flunking students) put in place many practices (e.g., extensive remedial programs) that they hope will alleviate the problems eventually. But at the time of the survey, principals saw no evidence that extensive remedial programs were making it possible for more students to earn promotion.

There are several other significant predictors of a school's retention rate in the middle grades. The grade-organization effects (col. 2, rows 1–5) indicate that the retention rate in middle schools is significantly lower than that found in 7–12, junior high, and 7–8 schools but is

TABLE 5

Percent of Seventh Graders Attending Schools That Offer Various Remedial Programs

	AVERAGE ACADEMIC ABILITY OF STUDENTS ON ENTRY			
REMEDIAL PROGRAMS	Considerably Below the National Norm ($N = 58$)	Near the National Norm ($N = 1{,}068$)	Considerably Above the National Norm ($N = 148$)	χ^2
Extra work or homework by classroom teacher	42	47	50	1.41
Pull-out program in reading or English	60	62	72	5.37*
Pull-out program in math	50	53	56	.92
Extra subject period instead of elective or exploratory course	24	28	34	2.68
After-school or before-school coaching sessions	69	57	73	15.44**
Saturday classes	0	3	4	1.83
Summer school	68	53	52	4.69*

---

* $P < .10$.

** $P < .01$.

TABLE 6

Standardized Regression Coefficients from Analyses Exploring the Antecedents and Consequences of the Number of Remedial Programs Offered

Effect	Number of Remedial Programs	Middle-Grades Retention Rate	Percent of Boys Who Will Probably Not Graduate from High School	Percent of Girls Who Will Probably Not Graduate from High School
Grade Organization:				
(1) K–8 vs. middle school	−.05	−.06*	−.06*	−.06*
(2) K–12 vs. middle school	−.04	.01	−.06*	−.05
(3) 7–12 vs. middle school	−.06*	.08*	−.11**	−.09**
(4) Junior high vs. middle school	−.04	.06*	−.03	−.02
(5) 7–8 vs. middle school	−.06	.07*	−.02	.01
Region:				
(6) West	.04	−.08*	.10**	.10**
(7) Midwest	−.04	.00	.01	.02
(8) South	.04	.11*	.06	.04
(9) Percent minority students	.06	.12**	.11**	.10**
(10) Percent families below poverty line	.00	.10*	.08*	.09*
(11) Percent professional families	.06	.02	−.13**	−.11**
(12) Average ability of students on entry	.08*	−.17**	−.28**	−.28**
(13) Population of SMSA	.10**	−.03	.06*	.09**
Number of courses a student can fail without being retained:				
(14) 0 vs. 3 or more	. . .	.09**	. . .	. . .
(15) 1 or 2 vs. 3 or more	. . .	.11**	. . .	. . .
(16) Number of remedial programs	. . .	.08**	−.02	−.02
(17) School's average retention rate across the middle grades	. . .	. . .	.09**	.08**
Adjusted R^2	.03	.17	.31	.29

NOTE.—Cell entries are standardized regression coefficients.
* $P < .05$
** $P < .01$

significantly higher than that found in K–8 schools. Retention rates are highest in the South and lowest in the West (rows 6–8). As might be expected, retention rates are higher in schools that serve many minority students, families living in poverty, and low-ability students (rows 9, 10, and 12). Finally, school policies concerning the number of courses students can fail and still be promoted also affect retention rates (e.g., schools that allow students to be promoted even if they fail three or more courses have lower retention rates than do other schools).

The sheer number of remedial programs offered does not affect principals' predictions of the percent of their current seventh graders who will not graduate from high school (cols. 3 and 4, row 16). There is, however, a significant positive effect of the average retention rate on estimated dropout rates (row 17). This finding is congruent with evidence from previous studies which suggests that holding students back "*increases* rather than decreases their risk of dropping out of school" (Grissom and Shepard 1989, p. 34).

The Extra-Subject-Period Approach to Remediation

Of the remedial practices included on the survey instrument, the practice of providing students who need extra help with an extra subject period during the school day (e.g., instead of an elective or exploratory course) seems especially promising. Remedial activities that occur outside of the regular school day—after-school or before-school coaching sessions, Saturday classes, or summer school—are often not well attended by the students who need the most extra help to master basic skills and pass courses. Including the "coaching class" as part of a low achiever's regular school day guarantees that more of the students who need help will actually receive it. Likewise, remedial programs using the extra-subject-period approach may be preferable to pull-out programs because students do not miss part of their other academic instruction (e.g., a student is not pulled out of social studies or science to receive extra help in reading or math), and being pulled out of class to receive help is a highly visible public event that increases the labeling and stigmatizing of low achievers. In contrast, fewer classmates may know or care that low achievers are receiving extra academic instruction during activity period rather than attending one of the other available electives, activities, or minicourses.

In many schools, students have two or more periods for elective subjects, so students who receive extra help with basic skills during one period are not excluded from exploring new subjects. Analyses reported elsewhere (Mac Iver and Epstein 1990, p. 24) indicate that schools that use the extra-subject-period approach to remediation have significantly more extensive exploratory programs than do other schools. That is, in these schools, even though some students devote some of their elective time to catching up, it is still the case that substantially greater proportions of students receive the opportunity to explore traditional electives (e.g., foreign language and home economics) and innovative minicourses in a variety of topics.

Regression analyses reveal that principals in schools that use the extra-subject-period approach to remediation do indeed report slightly lower expected dropout rates for both boys and girls (after controlling for all the variables in table 6 that are significant predictors of dropout rate). In schools that use this approach, the principal's estimates of the percentage of girls who will drop out is 1.4 percent below the rate that would otherwise be expected ($\beta = -.06$, $P = .02$). For boys, use of the extra-subject-period approach is associated with a lessening of the estimated dropout rate by 1.3 percent ($\beta = -.05$, $P = .04$). None of the other remedial practices in the questionnaire is significantly associated with principals' predictions concerning dropout rates. These approaches need further study to determine if and how they help students succeed.

Easing the Transition to a Middle Grades School

More than 88 percent of the public school students in the United States enter a new school as they make the transition to the middle grades (Epstein and Mac Iver 1990). There has been considerable concern about the negative effects that such school transitions can have on early adolescents (e.g., Blyth et al. 1983; Eccles and Midgley 1989; Eccles et al. 1984; Simmons and Blyth 1987). In response to this concern, many middle-grades schools have developed school transition programs (Epstein and Mac Iver 1990) and the National Middle School Association has begun officially recommending the use of such programs ("Resolutions" 1990).

Principals described the activities used with students, parents, and staff in their schools to ease the transition of students to the middle grades (see Appendix, variable XI). The three most common activities (used by over 40 percent of the principals) were having elementary school students visit the middle-grades school, having middle-grades and elementary administrators meet together on articulation and programs, and having middle-grades counselors meet with elementary counselors or staff.

Some potentially promising activities were infrequently used, perhaps because they are more difficult to implement. For example, only 20 percent or fewer of the principals indicated use of the following practices: having elementary school students attend regular classes at the middle-grades school, having summer meetings at the middle-grades school, and having a buddy program that pairs new students with older ones on entry to the school (Epstein and Mac Iver 1990).

Which types of middle-grades schools have the most extensive articulation and transition activities in preparing students for entry into their school? For these analyses, the measure of the extensiveness of the activities is the number of activities used at the time of the survey. This analysis excluded schools in which there was no transition to new buildings (e.g., K–8 and K–12 schools).

Articulation activities were significantly less extensive in 7–12 schools than in other schools that begin in the middle grades (see table 7, col. 1, row 1). Schools containing a large percentage of students living in poverty have less extensive articulation programs (row 8). Schools serving a large percentage of professional or managerial families, high-ability students, and populous urban areas have more extensive programs (rows 9–11).

There is evidence that an extensive articulation program may be beneficial. The standardized regression coefficient of .23 (in table 7, col. 2, row 12) indicates that principals in schools using numerous and diverse articulation activities are more likely to report that their articulation program is meeting student needs. Further, an extensive articulation program slightly—but significantly—increases the likelihood that students will succeed in their first year in the new school. That is, the −.07 in row 12 of column 3 indicates that fewer students are retained to repeat the transition grade in schools that have extensive transition programs. Of course, a school's retention policies (Appendix, variable X) also influence the percentage of students retained to repeat the transition grade (rows 13–15). A greater percentage of students are retained in schools where students are typically held back for failing one, two, or three courses or for excessive absence or lateness than in schools where students are not held back for these reasons (e.g., schools where students are held back only for failing four or more courses).

Discussion

How much do our data support the idea that middle-grades schools will be more successful if they adopt supportive structures, practices,

TABLE 7

Summary of Regression Analyses Exploring the Antecedents and Consequences of the Extensiveness of Articulation/Transition Activities for Students Making the Transition to the Middle Grades

Effect	Extensiveness of Articulation/ Transition Activities	Strength of Articulation/ Transition Program	Percent of Students Retained to Repeat the Transition Grade
Grade organization:			
(1) 7–12 vs. middle school	−.10**	.07*	.11**
(2) Junior high vs. middle school	−.01	.00	.08**
(3) 7–8 vs. middle school	−.02	.04	.09**
Region:			
(4) West vs. Northeast	.06	−.11**	−.08*
(5) Midwest vs. Northeast	−.05	−.04	−.03
(6) South vs. Northeast	−.01	.05	.12**
(7) Percent minority students	.06	.00	.19**
(8) Percent families below poverty line	−.17**	−.11**	.12**
(9) Percent professional families	.13**	−.01	.06
(10) Average ability of students on entry	.12**	.11**	−.18**
(11) Population of SMSA	.10**	.00	−.02
(12) Extensiveness of articulation/transition activities		.23**	−.07*
Major reasons students are retained:			
(13) Failing 1 course			.08**
(14) Failing 2 or 3 courses			.10**
(15) Excessive absence or lateness			.09**
Adjusted R^2	.13	.11	.26

NOTE.—Cell entries are standardized regression coefficients.
* $P < .05$.
** $P < .01$.

and services that leading educators in the middle-school movement often recommend as being especially responsive to the needs of early adolescents? Are there clear payoffs—measurable benefits to students or to the school program—if schools establish group advisory periods, interdisciplinary teams, provide remedial activities, and conduct extensive articulation practices?

First, it must be admitted that the "clear payoffs" question cannot be answered conclusively without information on student achievement, attitudes, or other important measures. Although principals' estimates of the strength of their middle-grades programs and of the benefits that result from responsive practices are important, they may or may not be related to student outcomes. Further, principals' estimates of future dropout rates and of current retention rates are undoubtedly imperfect reflections of actual dropout and retention rates. But it also is true that, in the context of existing research on responsive middle-grades practices, the data and analyses of this survey greatly extend knowledge of what practices are being implemented, the types of schools that are adopting or rejecting recommended practices, and the potential effects of these practices.

The evidence from principals suggest that most of the recommended practices yield measurable but modest benefits. For example, based on our data, a school in which the average frequency of occurrence of nine supportive group advisory activities is weekly rather than a few times per year is predicted to save 1 percent of the schools' students from dropping out before they finish high school. A school that provides an extra subject period within the school day to those students who need coaching or remediation is predicted to reduce its dropout rate by almost 1.5 percent. A school that uses the average number of articulation/transition practices is predicted to raise the percentage of students who succeed in their first year at the new school by approximately 1 percent over the promotion rates observed in otherwise similar schools that provide no special articulation/transition activities. Middle-grades programs in schools that balance a well-implemented interdisciplinary teacher team organization with a continuing commitment to departments are rated as much stronger by their principals (almost three-fourths of a point stronger on a 4-point scale) than are middle-grades programs in schools where teams, common planning periods, team leaders, and department heads are absent.

These results support the use of responsive practices and may understate their benefits. The potential benefits of responsive practices may be still greater than the average benefits reported here because some of the measures of practices were gross estimates of general aspects or broad distinctions in practices. For example, the measure of the extra-subject-period approach to remediation was a simple dichotomy. It distinguished schools using any variety of the extra-subject-period approach from schools that did not offer extra subject periods during the school day. Schools that provided intensive help during the extra subject period were lumped together with schools that provided little remedial instruction during the extra subject period (e.g., schools in which the period is more like a "study hall" than a "coaching period"). The benefits of having extra subject periods of intensive, well-organized, remedial instruction are undoubtedly larger than the average benefits of generic extra subject periods.

Further, the combined benefits of using several responsive practices simultaneously are larger than the benefits of using any one practice by itself. For example, schools implementing a strong group advisory program, an extra-subject-period approach to remediation, and responsive grading practices (Mac Iver, 1990) achieve more than a 3 percent reduction in expected dropout rates. Also, there are other likely benefits of responsive practices that were not measured at all in this study. For example, the typical cumulative effects of being in a responsive middle-grades school for three entire years on young adolescents' motivation to learn, achievement, and engagement and satisfaction with education may be substantial.

In this study (and in any study examining the relations between educational practices and outcomes), some of the observed relations may be spurious. We have controlled for a large number of possible "confounding variables" (e.g., average ability of students upon entry; percent of professional/managerial families in the school; percent of minority students; retention policies; regional differences in education policies; grade span; percent of families below the poverty line; urbanicity), but some important but less obvious variables may have been ignored. Thus our conclusions must be viewed as tentative, rather than as definitive. Still, the results of this study give justifiable encouragement to the many educators who have been calling and working for more responsive structures and services in the middle grades.

The results suggest, however, that to realize the benefits of a responsive practice, schools must make sure that practices are implemented properly. For example, a group advisory period will yield few benefits unless the teachers actually use the time to provide frequent social and emotional support activities to the students. Similarly, schools that organize their faculty into interdisciplinary teams without taking the steps necessary to make this organization work (e.g., establishing team leaders and common planning periods and training members how to use team planning effectively) may reap few benefits from teaming.

August 1991

Middle School Grades

A departmental organization and emphasis is not usually recommended by the leading educators in the middle-school movement. However, the data from principals suggest that schools that decide to emphasize departments and take the steps necessary to make this commitment work (establishing department heads, common planning periods for departments, and within-department team teaching) may be able to strengthen their programs just as much as schools that choose an interdisciplinary emphasis and take the steps necessary to make this emphasis work. What is equally important, we have seen that an interdisciplinary team organization and a departmental organization are not mutually exclusive.

Not all alternative approaches are equally beneficial, however. On the limited set of school-level outcomes examined in this study, the provision of an extra subject period during the school day was more beneficial than other approaches to remediation (presumably because of higher attendance and lower stigmatization of low achievers when the extra-period approach is used).

Educational researchers concerned with the middle grades are frequently asked, "What is the best grade span for a middle-grades school?" Overall, the responsive practices considered in this article are found most consistently in 6–8 middle schools. None of the other grade organizations used responsive practices significantly more than these schools. Still, grade organization is not a *consistent* determinant of responsive middle-grades practices. For example, although, on average, K–8 and K–12 schools are significantly less likely than 6–8 schools to implement interdisciplinary teaming, they are just as likely as 6–8 schools to use supportive activities during advisory group periods. Overall, 7–9 junior high and 7–12 schools use fewer responsive practices than other schools. But some junior high and 7–12 schools are as responsive as some middle schools on some practices.

One should not forget that the conclusions concerning the antecedents and consequences of responsive practices are based on data from public schools. As noted earlier, the sample did not include any Catholic or other private schools. Some of the effects described here (e.g., less use of interdisciplinary teaming in K–8 schools) may or may not generalize to private schools.

Many states and many school districts are attempting to restructure education in the middle grades. For example, 20 states have formed or are forming special task forces to examine the status of education in the middle grades and to make recommendations for improvement (Children's Defense Fund 1988; also see California State Department of Education [1987] and Maryland State Department of Education [1990]). Also, several major foundations (such as the Carnegie Corporation and the McConnell Clark Foundation) are attempting to stimulate restructuring efforts in the middle grades. It is, therefore, a critical time for building a solid middle-grades research base. For example, studies that explore the natural variation in middle-grades practices in the "real world" and that test the effects of these variations on a school's level of success are needed to assist educators in evaluating and selecting alternative approaches to middle-grades improvement.

Appendix
Variables Used in the Regression Analyses

This Appendix contains selected questionnaire items from *Education in the Middle Grades*, a national survey of practices and trends conducted in the spring of 1988 with a large, representative sample of public schools that include grade 7. The complete questionnaire may be found in Epstein and Mac Iver (1990). Throughout the questionnaire, principals were reminded to focus on their schools' practices "in the middle grades (in grades 5 through 9 for the grades that are in your school)."

Practices, Programs, Policies, and Staff in the Middle Grades

I. Use of Supportive Activities during Group Advisory Period

Each school was assigned a score representing the principal's mean response to the following set of items:

How frequently do the following activities occur during a homeroom or group advisory period in your school?

(*e*) Meet with individual students about problems.
(*f*) Give career information and guidance.
(*g*) Discuss academic problems or issues.
(*h*) Discuss personal or family problems.
(*i*) Discuss social relationships and peer groups.
(*j*) Discuss health issues, e.g. drug use prevention, family planning, etc.
(*k*) Discuss moral or ethical issues and values.
(*l*) Discuss intergroup relations and multi-cultural issues.
(*m*) Develop student self confidence and leadership.

The response scale for each item was: *Daily* (5), *Weekly* (4), *Monthly* (3), *A Few per Year* (2), and *Never* (1).

II. Guidance Counselors in the Middle Grades

About how many different students are assigned to each guidance counselor? (Give the guidance counselor-student ratio.)

If you have no guidance counselors, write none here: __________ and skip to the next question.

Middle School Grades

NUMBER OF STUDENTS PER COUNSELOR: ____________

NOTE.—Two variables were created based on responses to this question: A "students-per-counselor ratio" and a "presence-of-guidance-counselor" dummy variable (coded "1" for schools with guidance counselors and coded "0" otherwise).

III. Teacher Certification

Middle schools often have some mix of teachers trained and certified for the elementary grades or secondary grades or middle grades. How many of the teachers in your school are trained and certified (including provisionally certified teachers) in these different ways? (Please give your best estimates of these numbers.)

(*a*) Teachers with ELEMENTARY certification.

(*b*) Teachers with SECONDARY SUBJECT-MATTER certification.

(*c*) Teachers with specific MIDDLE GRADES certification (separate from elementary or secondary).

(*d*) UNCERTIFIED teachers.

(*e*) Other (describe).

NOTE.—Responses to this question were used to compute the percent of teachers with each type of certification.

IV. Implementation of an Interdisciplinary Team Approach

A school's emphasis on an interdisciplinary team approach to instruction and school organization was determined based on the principal's responses to 3 questions:

Is this practice part of your middle grades program now?

(*a*) Interdisciplinary teams of teachers who share the same students.

(*b*) Common planning period for members of interdisciplinary teams.

Does your school use INTERDISCIPLINARY Team Teaching? Two or more teachers of DIFFERENT SUBJECTS share the same group of students and/or coordinate their instructional program across subjects.

Circle all grades in which you use INTERDISCIPLINARY teams: 5 6 7 8 9

NOTE.—The "Implementation of an Interdisciplinary Team Approach" composite variable was equal to the number of yes responses on items (*a*) and (*b*) above, plus the proportion of grades in which interdisciplinary teams were used (maximum composite score = 3, minimum = 0).

V. Implementation of a Departmental Organization and Emphasis

A school's level of "commitment to departments" was determined based on the principal's responses to 3 questions:

Is this practice part of your middle grades program now?

(*a*) Departments organized with their own chairpersons or heads.

(*b*) Common planning period for members of departments.

Mac Iver and Epstein

Does your school use DEPARTMENT (SINGLE-SUBJECT) Team Teaching? Teachers in the SAME DEPARTMENT plan and teach together creating small group and large group activities by combining classes or regrouping students.
Circle all grades in which you use DEPARTMENT teams: 5 6 7 8 9
NOTE.—The "Commitment to Departments" composite variable was equal to the number of yes responses on items (*a*) and (*b*) above, plus the proportion of grades in which department teams were used (maximum composite score = 3, minimum = 0).

VI. Amount of Common Planning Time for Interdisciplinary Team Members

How much COMMON planning time is OFFICIALLY SCHEDULED EACH WEEK for the interdisciplinary team?
No official common planning time (1), Less than 30 minutes a week (2), Between one-half and 1 hour per week (3), Between 1 and 2 hours per week (4), Between 2 and 3 hours per week (5), More than 3 hours per week (6).

VII. Use of Common Planning Time on Interdisciplinary Teams

In a typical planning period for an interdisciplinary team, about how much time is spent on the following activities? Circle one choice for each activity that comes closest to your estimate of the work your teachers do during team planning meetings.
(*a*) Individual Teacher Preparation. Teachers work on their own lessons, tests, grades.
(*b*) Coordinate Content. Teachers decide common themes and related topics for instruction.
(*c*) Revise Schedules. Teachers arrange or alter schedules for classes that need more time.
(*d*) Regroup Students. Teachers arrange small or large groups of students to match lessons to abilities.
(*e*) Diagnose Individual Students. Teachers discuss problems of specific students and arrange help.
(*f*) Plan Special Events. Teachers arrange assemblies, trips, or other team activities.
(*g*) Conduct Conferences with Parents. Teachers meet as a team with individual parents to solve problems, provide assistance.
The response scale was:
How Much Time Per Planning Period?
None Little Less than half About half More than half
NOTE.—The "proportion of common planning time spent on team activities" composite variable was the mean of each principal's responses to (*b*)–(*g*).

VIII. Establishment of Formal Leaders for Interdisciplinary Teams

How is the leader chosen for the interdisciplinary team of teachers? (Circle one.)

August 1991

133

Middle School Grades

No leader is identified 1
Leader emerges informally as the team works together 2
Appointed by principal or other school official 3
Elected by other members of the teaching team 4
Leader rotates among members over time 5

NOTE.—The measure, "Have team leader?" was coded "0" if there were no team leaders or "informal" team leaders (options 1 or 2 above) and was coded "1" otherwise.

IX. Extensiveness of Remedial Instruction Activities

All schools have some students who fall behind or learn more slowly than other students. Does your school offer any of the following remedial activities for these students? (Circle all that apply.)

No special programs, it is up to students to stay on grade level 1
Extra work or homework by classroom teacher 2
Pull-out program in reading or English 3
Pull-out program in math 4
Extra subject period instead of elective or exploratory course 5
After-school or before-school classes or coaching sessions 6
Saturday classes 7
Summer school 8
Other (describe) 9

NOTE.—The extensiveness of a school's remedial instruction program was measured by counting the number of different programs offered by the school. Practices 1 or 2 ("No special programs" or "extra work or homework") were not included in this count.

X. Retention Policies/Major Reasons Students Are Retained

What are the major reasons most students are retained to repeat a grade in your school? (Circle all that apply as major reasons that students repeat the middle grades.)

Failing one course 1
Failing two or three courses 2
Failing more than three courses 3
Excessive absence or lateness 4
Failing achievement or proficiency tests 5
Other (describe) 6

XI. Organization of the Transition from the Elementary to the Middle Grades

How do you organize the transition from the ELEMENTARY to the MIDDLE grades? (Circle the numbers to the right of ALL of your present practices.)

No transition—middle grades continue in K–8 program 1

American Journal of Education

No special activities until students arrive in the fall	2
Middle grades students present information at elementary school	3
Elementary school students visit middle grades school for assembly	4
Elementary school students attend regular classes at middle grades school	5
Parents visit middle grades school while children are still in elementary school	6
Parents visit middle grades school for orientation in the fall after children have entered	7
Summer meetings at the middle grades school	8
Buddy or big brother/sister program pairs new student with older one on entry	9
Middle grades and elementary teachers meet together about courses and requirements	10
Middle grades and elementary administrators meet together on articulation and programs	11
Middle grades counselors meet with elementary school counselors or staff	12
Other (describe)	13

CHARACTERISTICS OF THE SCHOOL

XII. Grade Organization

What are the LOWEST and HIGHEST grade levels in your school? (Circle 2 choices.)

Pre-K K 1 2 3 4 5 6 7 8 9 10 11 12

XIII. Region (as defined by the U.S. Bureau of Census)

Schools were categorized by region:
West
Midwest
South
Northeast
This categorization was represented in the analyses by three dummy variables: the Northeast served as the control category.

XIV. Population of SMSA/Urbanicity

The population of the urbanized area of which the school is a part (in 100s). This includes the number of people living in the entire densely settled area around a city (e.g., people living in nearby suburbs or outlying cities and counties.) Schools in locations that are not in (nor adjacent to) an urbanized area are assigned a 0 on this variable.

August 1991

Middle School Grades

CHARACTERISTICS OF THE SCHOOL'S STUDENTS

XV. Percent Minority Students

Approximately what percentage of your present students are members of the following racial or ethnic groups?

(*a*) Black/Afro-American .... ______ %
(*b*) Hispanic-American .... ______ %
(*c*) Asian-American .... ______ %
(*d*) American Indian .... ______ %

XVI. Percent Families Below Poverty Line

The Orshansky Percentile for the school.

XVII. Percent Professional Families

Approximately what percentage of the students currently enrolled in your school are from families in the following categories?

(*a*) Professional and managerial personnel .... ______ %

XVIII. Average Ability of Students on Entry

How would you rate the average academic ability of students when they ENTER this school?

Considerably above the national norm (5), Somewhat above the national norm (4), At the national norm (3), Somewhat below the national norm (2), Considerably below the national norm (1).

OUTCOMES FOR THE SCHOOL AND ITS STUDENTS

XIX. Strength of the School's Guidance Programs

How well do your PRESENT practices match your IDEAL program for guidance, advice, and counseling of students in the middle grades?

EXCELLENT—present practices fit students' needs exactly (4), GOOD—basic practices are in place, minor changes needed (3), FAIR—need to improve or add several practices (2), WEAK—need to design new practices, major changes needed (1)

American Journal of Education

138

XX. Percentage of Boys and Percentage of Girls Who Probably Will Not Graduate from High School

Based on your experience, past records, or best guesses, please estimate the percent of your present seventh grade BOYS and GIRLS who will PROBABLY NOT graduate from high school.

(*a*) percent of present seventh grade boys who will probably NOT graduate from high school .... ______ %

(*b*) percent of present seventh grade girls who will probably NOT graduate from high school .... ______ %

XXI. Strength of the School's Overall Middle Grades Program

How well do your present practices match your IDEAL of a successful program for students in the middle grades?

EXCELLENT—present practices fit students' needs exactly—exemplary program (4), GOOD—basic practices are in place, minor changes needed—solid program (3), FAIR—need to improve or add some practices—developing program (2), WEAK—need to design new practices and major revisions—changing program (1)

XXII. Benefits Resulting from Use of Interdisciplinary Teams

Each school was assigned a score representing the principal's mean response across four different types of benefits:

There are potential benefits in using interdisciplinary teams in the middle grades. How often do you think the following occur as a result of interdisciplinary teams in your school?

Students identify with the team, build team spirit, and improve school work and attitudes (1 = Never, 5 = Always).

Individual student problems are recognized quickly and solved effectively (1 = Never, 5 = Always).

Teachers use other team members as sources of social support and understanding (1 = Never, 5 = Always).

Instruction is more effective due to integration and coordination across subjects and courses (1 = Never, 5 = Always).

XXIII. Strength of the School's Transition/Articulation Program

How well do your present practices match your IDEAL program for students' smooth transitions to and from the middle grades? Circle one choice.

EXCELLENT—present practices fit students' needs exactly (4), GOOD—basic practices are in place, minor changes needed (3), FAIR—need to improve or add several practices (2), WEAK—need to design new practices and major changes (1).

XXIV. Retention Rates in the Middle Grades

At the end of last school year (after summer school), about how many students were promoted to the next grade and how many were retained to repeat the same grade this year? (Give approximate numbers.)

For 1987 School Year after Summer School	NUMBER OF STUDENTS	
(*a*) From Grade 5	_ promoted	_ retained
(*b*) From Grade 6	_ promoted	_ retained
(*c*) From Grade 7	_ promoted	_ retained
(*d*) From Grade 8	_ promoted	_ retained
(*e*) From Grade 9	_ promoted	_ retained

Notes

We are indebted to colleagues at CDS who collaborated in the design of the questionnaire for this survey, including Henry Jay Becker, Jomills H. Braddock III, and James M. McPartland. We appreciate the support of the National Association of Elementary School Principals and the National Association of Secondary School Principals in the survey effort. We are especially grateful to the principals in schools that contain grade 7 who invested their time to provide information about middle-grades education. We would also like to thank Allan Wigfield and anonymous reviewers for helpful comments on an earlier draft. This research is sponsored by the Office of Educational Research and Improvement, U.S. Department of Education, under grant number OERI-G-86-90006. The opinions presented are those of the authors and no official agreement by OERI should be inferred.

References

Alexander, W. M., and P. S. George. *The Exemplary Middle School*. New York: Wiley, 1981.

Briggs, T. H. "Extra-Curricular Activities in the Junior High School." *Educational Administration and Supervision* 8 (1922): 1–9.

Blyth, D. A., R. G. Simmons, and S. Carlton-Ford. "The Adjustment of Early Adolescents to School Transitions." *Journal of Early Adolescence* 3 (1983): 105–20.

California State Department of Education. *Caught in the Middle: Educational Reform for Young Adolescents in California Public Schools*. Sacramento, Calif.: California State Department of Education, 1987.

Carnegie Task Force on the Education of Young Adolescents. *Turning Points: Preparing American Youth for the 21st Century*. New York: Carnegie Corporation, 1989.

Children's Defense Fund. *Survey of State Policies and Programs for the Middle Grades*. Washington, D.C.: Children's Defense Fund, 1988.

Eccles, J. S., and C. Midgley. "Stage/Environment Fit: Developmentally Appropriate Classrooms for Early Adolescents." In *Research on Motivation in Education*, vol. 3, edited by R. E. Ames and C. Ames. New York: Academic Press, 1989.

Eccles, J. S., C. Midgley, and T. Adler. "Grade-related Changes in the School Environment: Effects on Achievement Motivation." In *The Development of Achievement Motivation*, edited by J. G. Nicholls. Greenwich, Conn.: JAI, 1984.

Epstein, J. L. *Schools in the Center*. Baltimore: CREMS, 1988.

Epstein, J. L., and D. J. Mac Iver. *Education in the Middle Grades: National Practices and Trends*. Columbus, Ohio: National Middle School Association, 1990.

Grissom, J. B., and L. A. Shepard. "Repeating and Dropping Out of School." In *Flunking Grades: Research and Policies on Retention*, edited by L. A. Shepard and M. L. Smith. Philadelphia: Falmer, 1989.

Haffner, A., S. Ingels, B. Schneider, D. Stevenson, and J. A. Owings. *A Profile of the American Eighth-Grader: NELS:88 Student Descriptive Summary*. Washington, D.C.: National Center for Education Statistics, 1990.

Hall, G. S. *Adolescence: Its Psychology and Its Relations to Physiology, Anthropology, Sociology, Sex, Crime, Religion, and Education*. New York: Appleton Century, 1905.

Hieronimus, N. C. "The Teacher-Advisor in the Junior High School." *Educational Administration and Supervision* 3 (1917): 91–95.

Jackson, A. W., and D. W. Hornbeck. "Educating Young Adolescents: Why We Must Restructure Middle Grade Schools." *American Psychologist* 44 (1989): 831–36.

Kitson, H. D. "Extra-curricular Activities as a Means of Guidance." *Vocational Guidance Magazine* 4 (1926): 356–61.

Koos, L. V. *The Junior High School*. Boston: Ginn, 1927.

Mac Iver, D. J. *A National Description of Report Card Entries in the Middle Grades*. CDS Report 9. Baltimore: Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students, 1990.

Mac Iver, D. J., and J. L. Epstein. *How Equal Are Opportunities for Learning in Disadvantaged and Advantaged Middle Grades Schools?* CDS Report 7. Baltimore: Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students, 1990.

Mac Iver, D. J., and J. L. Epstein. "Middle Grades Education: Middle Schools and Junior High Schools." In *Encyclopedia of Educational Research*, 6th ed., edited by M. Alkin. New York: Macmillan, in press.

Maryland State Department of Education. *What Matters in the Middle Grades: Recommendations for Maryland Middle Grades Education*. Report of the Maryland Task Force on the Middle Learning Years. Baltimore: Maryland State Department of Education, 1990.

Merenbloom, E. Y. *The Team Process in the Middle School: A Handbook for Teachers*. Columbus, Ohio: National Middle School Association, 1986.

Perlstein, D., and W. Tobin. *The History of the Junior High School: A Study of Conflicting Aims and Institutional Patterns*. Paper commissioned by the Task Force on Education of Young Adolescents. New York: Carnegie Council on Adolescent Development, 1988.

August 1991

Middle School Grades

"Resolutions 1989–1990." *Middle Ground* 17 (Spring 1990): 5.
Simmons, R. G., and D. A. Blyth. *Moving into Adolescence: The Impact of Pubertal Change and School Context*. Hawthorn, N.Y.: Aldine de Gruyter, 1987.
Smith, W. A. *The Junior High School*. New York: Macmillan, 1925.
Van Hoose, J., and D. Strahan. *Young Adolescent Development and School Practices: Promoting Harmony*. Columbus, Ohio: National Middle School Association, 1988.
Vars, G. F. *Interdisciplinary Teaching in the Middle Grades: Why and How*. Columbus, Ohio: National Middle School Association, 1987.

American Journal of Education

Educational Evaluation and Policy Analysis
Summer 1990, Vol. 12, No. 2, pp. 157–173

School Competency Testing Reforms and Student Achievement: Exploring a National Perspective

Linda F. Winfield
The Johns Hopkins University Center for Research on Effective Schooling for Disadvantaged Students

This study investigates the relationship between school-level minimum competency testing (MCT) programs and student reading proficiency as measured by the 1983–1984 National Assessment of Educational Progress (NAEP). Comparisons of student-level proficiency outcomes within race/ethnic groups (White, Black, and Hispanic) were made after adjusting for individual and school-level variables for the 4th-, 8th-, and 11th- grade NAEP samples. In general, results indicated a higher level of proficiency among students in Grades 8 and 11 attending schools with MCT programs compared with their counterparts in schools without such programs. No advantage of attending such schools was identified for students in Grade 4.

Since the early 1970s and throughout the 1980s, numerous reform initiatives have sought to increase the accountability and effectiveness of public education in America. Timar and Kirp (1989) note: "Since 1983 the states have generated more rules and regulations about all aspects of education than in the previous twenty years" (p. 506). Efforts to improve student achievement outcomes have included increasing graduation requirements and implementing assessment programs that define both standards of performance for students and standards of accountability for the educational system. Competency-based testing programs implemented at the local and state levels have continually increased as a primary method of reform (Airasian, 1987). Since the mid-seventies, over 35 states have required local school districts to give minimum competency tests (MCT) to students in elementary, junior high, or senior high school (Pipho, 1983). In 1984, 40 states were actively involved in some aspect of minimum competency testing, 19 states were using test performance as a basis for high school graduation, and 5 states were using tests as a basis for grade promotion (Anderson & Pipho, 1984; Pipho & Hadley, 1984).

The majority of MCT programs focus on improvement in basic skills in reading and math (Educational Commission of the States, 1984). One basic premise for implementing such programs is that the tests clearly specify learning objectives and encourage schools and teachers to focus instruction more precisely. Additionally, MCT

A version of this paper was presented at the annual meeting of the 1987 American Educational Research Association, Washington, DC. This research was supported by funding provided by the Educational Testing Service, Princeton, NJ, while the author was a National Assessment of Educational Progress Visiting Scholar, 1985–1986. Support for the preparation of this article was provided by the Center for Research on Effective Schooling for Disadvantaged Students, The Johns Hopkins University. The views expressed are those of the author and do not reflect the position of the supporting agencies.

Copyright 1990 by the American Educational Research Association. Reprinted by permission of the publisher.

results can be used as a basis for diagnosis and remediation of academic skills (Cohen & Haney, 1980) and consequently lead to higher student achievement. A major objection to MCT programs, however, is that they lead to "teaching to the test" and a narrow focus of instruction that neglects those skills not included on the tests (Broudy, 1980; Koretz, 1988). If this is the case, the overall quality of school programs is being adversely affected. Others suggest that competency tests and standards function more as symbolic and political gestures than as instrumental reforms (Ellwein, Glass, & Smith, 1988).

The Impact of MCT Programs

Two related areas of research on student outcomes provide a framework for conceptualizing the potential impact of competency-based programs. First, research on schools that are considered "unusually effective" in facilitating reading achievement in minority and low-socioeconomic (SES) populations suggests that MCT programs contribute to school-wide success through a clear definition of learning objectives, curriculum organization, and careful monitoring (Edmonds, 1979; Eubanks & Levine, 1983; Good & Brophy, 1986; Kyle, 1985; Mackenzie, 1983; Purkey & Smith, 1983, 1985; Stedman, 1987; Stringfield & Teddlie, 1988). In contrast, this research also indicates that school-wide improvement in student achievement is more likely to occur from increasing local decision-making and school-site responsibility. Change can be successfully implemented at the individual school building level, given the appropriate conditions, procedures, and support systems (Darling-Hammond & Wise, 1985; Fullan, 1985; Goodlad, 1984; Sirotnik & Clark, 1988; Sirotnik & Oakes, 1986). Thus, a combination of top-down and bottom-up approaches in implementing school improvement efforts appears to be necessary for success.

The second related area of literature, generally known as "school effectiveness" research, attempts to isolate the relative effect of specific school characteristics as compared with that of family background and SES on academic achievement. (See Madaus, Airasian, & Kellaghan, 1980, for a review of the evidence of this area). Some studies in this area, derived from a sociological perspective, indicate that student background characteristics have a higher impact on achievement than do school characteristics (Coleman et al., 1966; Jencks et al., 1972; Mosteller & Moynihan, 1972; Smith, 1972). Other studies, however, emphasize the combined effect of home and school factors (Mayeske et al., 1972; Mayeske, Okada, Cohen, Beaton, & Wisler, 1973). More recent studies, including those on "effective" schools, have provided additional evidence which indicates that specific school characteristics can facilitate achievement (Barr & Dreeben, 1983; Edmonds, 1979; Frederiksen, 1975; Good & Brophy, 1986; Kyle, 1985; Purkey & Smith, 1983, 1985; Stedman, 1987; Venezky & Winfield, 1979). Specific school level policies and institutional practices (e.g., grouping, instructional pace, and content coverage) are related to student reading achievement outcomes (Barr & Dreeben, 1983). The framework of the "nested layers" in which schools operate further suggests that actions at higher layers (e.g. district and state) influence conditions occurring at the school and classroom levels (Purkey & Smith, 1983). Considered in this manner, MCT programs may be viewed as a state, district, or school-level variable influencing academic achievement.

As a result of competency programs, increased attention has focused on outcome measures (Murphy, 1989), and more resources are targeted for students who need remediation. However, Black and low-SES students fail MCTs in substantially higher proportions than do White and higher-SES students (Jaeger, 1982; Jonas & Wallace, 1986; Linn, Madaus, & Pedulla, 1982; Serow, 1984), and remediation for students who had failed competency tests in reading was found to be less effective for Black students than for White students (Serow & Davies, 1982). Serow (1984) reported that in four states in which MCT programs had been implemented at the secondary level, Black students had a substantially lower passing rate than did Whites. Moreover, in one state, lower SES students were about one third less likely to pass the exam on their first attempt compared with higher SES students. Unfortunately, too few studies have assessed the impact on minority and low-SES students. In general, minority issues receive the most attention during test construction and validation (Ellwein et al., 1988).

Over 400 articles were published between 1977 and 1987 regarding MCT; however, 62% of this literature was rhetorical (Ellwein, Popp, & Neimann, 1988). There have been few empirical research studies of the effects of MCT on student achievement. Most studies have been conducted at the local or state level and have used as an outcome measure the percentage of students passing the test. Thus, if MCT programs enhance initial test-taking scores, then a decline in the percentage of students previously failing the test is taken as indirect evidence of improved student outcomes. However, what appears to be an indicator of improvement in students' basic skills might reflect either practice or regression to the mean (Serow, 1984). Additionally, the criteria for passing may fluctuate over time. One of the few studies that used standardized achievement as an outcome found that after implementation of MCT programs, ninth-grade math basic skills increased but did not show continual increases over the succeeding 3 years (Mangino & Babcock, 1986).

Despite the limited knowledge base on implementation effects, the number of states and districts that implemented testing reforms increased dramatically during the past decade. Some analysts suggested that these reforms will yield few returns because they are built upon and reinforce existing organizational arrangements (Chubb, 1988). Others suggest that the reforms may be successful because key organizational linkages in existing school structures have been tightened (Murphy, 1989). What impact have testing reforms had on student achievement outcomes nationally? Because of the variation among testing programs in contexts, procedures, and criteria for passing, it is difficult to compare student achievement outcomes on anything but an intrastate basis. The need for accurate information regarding school quality and educational reforms has been documented (Alexander-James, 1987). The National Assessment of Educational Progress (NAEP) will be redesigned to provide state-by-state comparisons in the area of reading and mathematics in 1992. Although there is evidence that basic reading skills improved nationwide over the last several years (NAEP, 1985), one analysis suggested that this trend could not be attributed to competency-testing programs because the upturn in achievement had already been under way a few years prior to the major growth of MCT reforms in the late 1970s (Congressional Budget Office, 1987). However, the aggregated trend data do not provide direct evidence of the effect of these programs on schools and students. The major substantive questions of interest in this study are two: (a) What is the relationship between school-level MCT programs and student reading achievement outcomes? (b) Does this relationship differ for various race/ethnic groups? A second concern is the feasibility and utility of using the new, redesigned NAEP (Messick, Beaton, & Lord, 1983) data in addressing these issues. For this reason, the study should be considered as exploratory in nature.

Method

Sample

The data for this study are from the 1983–1984 NAEP. The NAEP is funded by the Office for Educational Research and Improvement and is under a grant for the Educational Testing Service. Each NAEP assessment involves a random cross-sectional survey of in-school 9-, 13-, and 17-year olds. In the 1983–1984 assessment, in addition to sampling by age, Grades 4, 8, and 11 were also sampled. Each age/grade cohort included approximately 30,000 students. The 1983–1984 NAEP sample was based on a stratified, four-stage probability sampling design in which counties, schools, type of sessions, and students were sampled. In selecting schools, those in large cities with high concentrations of low-SES students and those in extremely rural areas were sampled at twice the rate of other schools. A total of 64 first-stage units was included in the sample to represent the 50 states and the District of Columbia, and assessments were conducted at 1,465 schools.

In addition to the assessment of student-level data, NAEP collected school-level information from school administrators. A principal or his or her representative completed the five-page questionnaire concerning staffing patterns, curriculum, and student services. The overall survey response rates were 81% for Grade 4, 75% for Grade 8, and 75% for Grade 11.

Subsample

Schools included in this study are a nonrandom subsample of the original NAEP sample. Schools were included only if the principal (a) responded to the school questionnaire and (b) provided responses to the minimum competency questions that were consistent across the items included. The school response rates for the item requesting information on minimum competency testing were lower than overall survey response rates and were 49% for Grade 4, 52% for Grade 8, and 60% for Grade 11. Analyses of schools that did not respond to the MCT item indicated no significant differences on the school- and student-level variables included in the study. These data and other information characterizing the schools and students in the study can be found in a discussion by Winfield (1987b). Because NAEP produces a representative national sample, each student or school has an associated sampling weight to account for the differential probability of selection and adjustments for nonresponse and poststratification. To ensure adequate representation, certain subgroups were sampled at a higher rate than the rest of the population. Thus, sampling weights were used in all analyses. (These weights were rescaled so that the sum of the weights equaled the number of cases included in each analysis. See NAEP, 1986a, 1986b, for procedures to be used when analyzing NAEP data.) The unweighted and adjusted weighted frequencies of schools and students in the total subsample are shown in Table 1.

In subsequent analyses, the size of the subsamples for Grades 4, 8, and 11 were 10,367, 10,829, and 13,513, respectively. These numbers represent 39.8% of the total NAEP Grade 4 cohort, 41.8% of the Grade 8 cohort, and 55.2% of the Grade 11 cohort. The number in each racial/ethnic group was 7,491 Whites, 1,733 Blacks, and 1,143 Hispanics in Grade 4; 7,574 Whites, 1,906 Blacks, and 1,349 Hispanics in Grade 8; and 9,203 Whites, 2,112 Blacks, and 2,198 Hispanics in Grade 11.

Reading Proficiency Outcome Variable

The goal of NAEP is to estimate group means rather than individual proficiency; thus, each respondent may answer only a subset of the total number of assessment items. In the 1983–1984 assessment, NAEP used a balanced incomplete block (BIB) spiraling procedure in which the total assessment was divided into blocks of 15 minutes each. Each student was administered three 15-minute blocks of items and a 6-minute block of general background questions. The BIB part of the method assigned blocks to booklets in such a way that each block appeared in the same number of booklets and each pair of blocks appeared in at least one booklet. The spiraling part of the method then cycled the booklets for administration so that no two students in any assessment session in a school received the same booklet. At each age group, each block is administered to approximately 2,000 students and each pair of blocks to a smaller number, depending upon the particular BIB design (NAEP, 1986a).

Item response theory (IRT) technology was used to estimate reading proficiency levels. This theory defines a student's probability of answering an item correctly as a mathematical function of an underlying proficiency or skill. Indicators of proficiency are computed as random draws from the expected distribution of proficiency of each respondent given the observed data, in this

TABLE 1
Unweighted and adjusted weighted frequencies of schools and students by school response to NAEP MCT item

Grade and response	Schools Unweighted	Schools Weighted	%	Students Unweighted	Students Weighted	%
Grade 4 min. comp.						
Yes	169	334	25.3	7.226	11.121	29.0
No	154	312	23.6	5.289	7.952	20.7
No response	340	676	51.1	13.489	19.222	50.3
Total	663	1.322	100	26.004	38.295	100
Grade 8 min. comp.						
Yes	141	262	27.0	6.744	13.426	32.1
No	110	235	24.3	4.521	8.431	20.1
No response	235	472	48.7	10.573	20.004	47.8
Total	486	969	100	21.838	41.921	100
Grade 11 min. comp.						
Yes	118	203	30.6	9.170	17.621	41.5
No	82	152	25.2	5.454	9.778	23.0
No response	131	238	44.2	8.119	15.067	35.5
Total	331	638	100	22.788	42.466	100

Note. NAEP = National Assessment of Educational Progress: MCT = Minimum Competency Test: min. comp. = minimum competency.

instance responses to NAEP reading exercises and background variables. (See Mislevy, 1985. for the statistical foundations of this approach.) The distribution of such draws. one taken for each respondent and weighted in inverse proportion to the respondent's probability of appearing in the sample. estimates the distribution of proficiency in the population as a whole or in a given subpopulation. Because the resulting indicators do not represent precise estimates of proficiency for individual examinees. five "plausible values" from this distribution are provided for each student. who was administered at least one block with reading items. The NAEP reading proficiency scale ranges from 0 to 500 with a mean of 305 and a standard deviation of 50.

School-Level. Individual. and Control Variables

Although a number of school-level variables were included in initial analyses (e.g., teacher turnover. hours of in-service. whether there was a Chapter I program. and student absenteeism). many of these were deleted in the final analyses because of multicolinearity. Variables included were identified through a combination of stepwise multiple regression and were based on theoretical relevance. Variables selected at the student level were family background. student academic behaviors. age. and sex. Family background consisted of responses to items on parental education. reading materials in the home. and the extent of family reading. Student academic behaviors consisted of students' responses to items requesting the number of pages read for school and the amount of homework. Control variables were region of the country. percentage of students on free lunch. and school racial composition. and school district SES. Two potential explanatory variables were school-level aggregate of instructional dollars per pupil and the presence/absence of a remedial program for students failing the MCT.

School-level effect of interest—(MCT) program in reading. Two of the four items on the NAEP questionnaire. taken from principal's self-reports. were used to identify schools implementing MCT programs. One item read: "In which of the following subjects are students required to pass a minimum competency test?" Respondents were required to answer either yes or no to items identifying several subject areas, one of which was reading. The second item read: "In what year was each of the following minimum competency tests first administered?" Responses to this item ranged from year 1976 to 1984. Affirmative responses to the MCT item on reading were then coded according to whether the program had been implemented prior to 1980. School- and student-level descriptive data for each grade level by school type and for each race/ethnic group included in the study are presented in Tables 2, 3, and 4.

Unadjusted reading proficiency means and standard deviations by grade, race/ethnic group, and school type are shown in Table 5.

Design and Data Analysis

What is the relationship between school-level MCT and student reading proficiency? In this multilevel analysis of schools, individual student reading proficiencies rather than school-level aggregate achievement (Burstein & Miller, 1981) were used as dependent measures. From a statistical framework, one might ask, What is the effect on student proficiency of having an MCT program in reading after controlling for regional variation, school-level SES, and individual student variables? Is the effect of MCT on reading proficiency outcomes the same for all race/ethnic groups? Ideally, to answer either of these questions, individual student-level achievement within schools must be examined for changes in the distribution of reading proficiency for various race/ethnic and SES groups over a period of time. However, NAEP data are cross-sectional and available for one time period only; thus this investigation is limited to the direction and strength of correlates of achievement.[1]

An analysis of covariance within a regression framework was conducted for each race/ethnic group (White, Black, and Hispanic) with each of the three grade cohorts.

TABLE 2
School- and student-level characteristics: Weighted averages by race/ethnic group by school type—Grade 4

	White MCT		Black MCT		Hispanic MCT	
Variable	Yes	No	Yes	No	Yes	No
School level						
% White students						
M	77	91	35	37	49	62
SD	21	12	30	38	27	36
% of students/free lunch						
M	39	39	64	62	46	48
SD	32	34	32	30	33	37
Instructional $ per pupil						
M	58	49	58	63	60	50
SD	23	23	22	15	22	28
Student level						
Family background[a]						
M	6.07	6.21	5.71	5.82	5.42	5.47
SD	2.14	2.02	2.17	2.10	2.28	2.11
Students' academic behaviors						
M	3.44	3.34	3.46	3.26	3.36	3.21
SD	1.70	1.67	1.75	1.80	1.62	1.61
Student's age						
M	9.25	9.29	9.34	9.45	9.38	9.36
SD	.52	.51	.61	.67	.65	.61

Note. MCT = Minimum Competency Test.
[a] Composite of parents' education plus possessions in the home.

TABLE 3

School- and student-level characteristics: Weighted averages by race/ethnic group by school type—Grade 8

	White MCT		Black MCT		Hispanic MCT	
Variable	Yes	No	Yes	No	Yes	No
School level						
% White students						
M	76	89	37	43	39	39
SD	22	16	31	34	27	26
% of students/free lunch						
M	23	41	49	58	44	61
SD	24	30	28	33	27	30
Instructional $ per pupil						
M	63	53	65	55	66	48
SD	15	13	16	11	14	11
Student level						
Family background[a]						
M	7.23	7.03	6.39	6.27	5.70	5.24
SD	1.62	1.57	1.87	1.95	2.03	2.04
Students' academic behaviors						
M	3.63	3.58	3.53	3.48	3.56	3.33
SD	1.90	1.68	1.75	1.70	1.77	1.70
Students' age						
M	13.3	13.3	13.4	13.7	13.4	13.6
SD	.52	.53	.73	.90	.57	.70

Note. MCT = Minimum Competency Test.

[a] Composite of parents' education plus possessions in the home.

This model tests the assumption that the within-group regression coefficients are homogeneous for schools with and without MCT programs and that one may test differences between groups after adjusting for the effects of other attributes. There were nine parallel regression equations, each incorporating the same predictor variables. Individual reading proficiencies were used as dependent measures. In accordance with suggested NAEP procedures, each "plausible value" was used as a dependent measure in a regression analysis. Thus, there were five regression equations for each of the nine race/grade groups, or 45 regression equations. The five resulting regression coefficients for MCT for each race/grade were then averaged to arrive at the reported effect. Standard errors were adjusted to reflect the variance due to uncertainty in these values and due to sampling.

Covariates at the student level were age, sex, family background, and students' academic behaviors. Covariates included at the school level were region of the country, percentage of students on free lunch, and school racial composition. The school-level variables for MCT and remedial program were each dummy coded—1 = yes and 0 = no—and remedial program and instructional dollars per pupil were included as two potential explanatory variables.

The interactions between the MCT variable and student age, school-level SES, region, MCT, percentage of students on free lunch, percentage of White students, family background, and students' academic behaviors were tested as a block, entered last, and were found to be nonsignificant. All regression analyses were conducted on students in the grade samples rather than age samples. Listwise deletion of missing cases was used in all analyses. Because the NAEP sample design employs stratifications and clustering (students within schools, schools within primary sampling units), the resulting sample has different statistical characteristics from those of a simple random sample. To ac-

TABLE 4
School- and student-level characteristics: Weighted averages by race/ethnic group by school type—Grade 11

Variable	White MCT		Black MCT		Hispanic MCT	
	Yes	No	Yes	No	Yes	No
School level						
% White students						
M	79	88	45	62	35	55
SD	19	15	32	21	29	31
% of students/free lunch						
M	15	22	38	39	34	40
SD	16	20	27	25	23	25
Instructional $ per pupil						
M	61	57	65	50	62	59
SD	17	14	15	19	16	11
Student level						
Family background[a]						
M	7.60	7.25	6.72	6.40	5.91	5.83
SD	1.47	1.41	1.66	1.62	1.87	1.71
Students' academic behaviors						
M	4.30	3.93	4.12	3.86	3.93	3.72
SD	2.06	1.94	1.91	1.94	1.93	1.93
Students' age						
M	17.0	17.0	17.3	17.3	17.3	17.4
SD	.60	.50	.80	.80	.80	.70

Note. MCT = Minimum Competency Test.

[a] Composite of parents' education plus possessions in the home.

TABLE 5
Unadjusted average reading proficiency by grade by race/ethnic group by school type

Race/ethnic group	Grade 4 MCT			Grade 8 MCT			Grade 11 MCT		
	Yes	No	No resp.	Yes	No	No resp.	Yes	No	No resp.
White									
M	220.0	227.0	219.6	267.7	260.6	266.9	298.1	290.8	291.1
SD	31.5	30.5	32.7	27.2	28.5	27.7	29.9	31.5	32.3
Black									
M	194.9	194.4	189.0	243.9	232.3	239.5	267.7	262.7	261.6
SD	28.3	28.6	30.5	27.2	27.8	28.0	28.2	31.5	31.3
Hispanic									
M	197.9	202.3	194.9	243.7	236.4	244.1	269.0	261.9	263.8
SD	31.1	29.8	32.7	29.8	28.5	27.2	32.2	34.4	34.0

Note. MCT = Minimum Competency Test; resp. = response.

count approximately for the effects of the sample design, a design effect of two was used. Kish and Frankel (1974) suggest that design effects for complex statistics from complex samples are greater than 1. This has the effect of dividing the sample size in half and using the adjusted sample size in the computation of errors. This method was used in lieu of the Educational Testing Service jacknife technique employed in estimating sampling variability of statistics included in official NAEP reports. (See Johnson, 1987, for a discussion of design effects used to adjust error estimates when using NAEP data.)

Results

The covariate-adjusted contrasts between MCT and non-MCT schools resulting from the regressions of reading proficiencies on school- and student-level variables for each race/grade cohort are presented in Table 6. For the sake of brevity, the term *effect* is used in discussing the relationships identified between MCT programs and achievement, although the findings are correlational and not necessarily causal. The first column presents the effect of MCT after adjusting for sex, age, region of the country, school-level SES, family background, and students' academic behaviors. The second column presents the effect after adjusting for all of the previously mentioned variables in addition to the explanatory variables, per pupil instructional dollars, and school-level remedial reading program.

Effects adjusted for all variables included in the regression analyses for each race/ethnic group by grade level are depicted in Figure 1. The complete regression equation for each grade group is shown in the Appendix.

Grade 4

At Grade 4, after controlling for all variables, there were no significant effects attributed to the MCT dummy variable for any of the race/ethnic groups.

Grade 8

At Grade 8, after adjusting for student and school-level variables, there were positive effects for both White and Black eighth graders. This effect represented about an 8-point (.29 *SD*) advantage for Whites and a 10-point (.38 *SD*) advantage for Blacks in average reading proficiency as compared with their respective counterparts in schools without MCT programs. Effect sizes indicated are calculated as the difference between treatment and comparison adjusted means divided by the standard deviation of the comparison group (Glass, 1977). The inclusion of remedial program and instructional dollars as variables explained part but not all of the MCT effect. Inclusion of these variables reduced the effect for White students by about 29% and for Black students by about 31%. No significant effect was isolated for Hispanic students.

Grade 11

At Grade 11, after controlling for student- and school-level variables, there were positive effects for all race/ethnic groups. This effect reflected a 2-point (0.6 *SD*) advantage in average reading proficiency for White students attending schools with MCT programs, a 7-point (.26 *SD*) advantage for Blacks, and a 6-point (.19 *SD*) advantage for Hispanics as compared with their respective counterparts in schools without MCT programs. Inclusion of remedial program and instructional dollars explained the effect for

TABLE 6
Covariate-adjusted contrasts for MCT (Minimum Competency Test) and non-MCT schools

Race/ethnic group	Grade 4		Grade 8		Grade 11	
	Adjusted[a]	Final step[b]	Adjusted	Final step	Adjusted	Final step
White						
b	−1.66	−1.08	7.79**	5.54**	2.18*	.42*
SE	1.09	1.08	1.29	1.51	1.01	1.58
Black						
b	0.30	2.89	10.90***	7.60***	6.62***	12.34***
SE	2.58	3.97	2.51	3.23	2.91	3.48
Hispanic						
b	−2.86	3.00	−1.21	0.06	5.93**	5.76**
SE	3.01	3.23	3.50	3.71	3.48	4.33

[a] At each grade level, effect is adjusted for sex, student age, regions of the country, school-level socioeconomic status, family background, and student academic behaviors.

[b] At each grade level, effect is adjusted for student- and school-level variables in addition to school-level remedial program (dummy coded 1 = yes, 0 = no) and instructional dollars per pupil.

* $p < .05$. ** $p < .01$. *** $p < .001$.

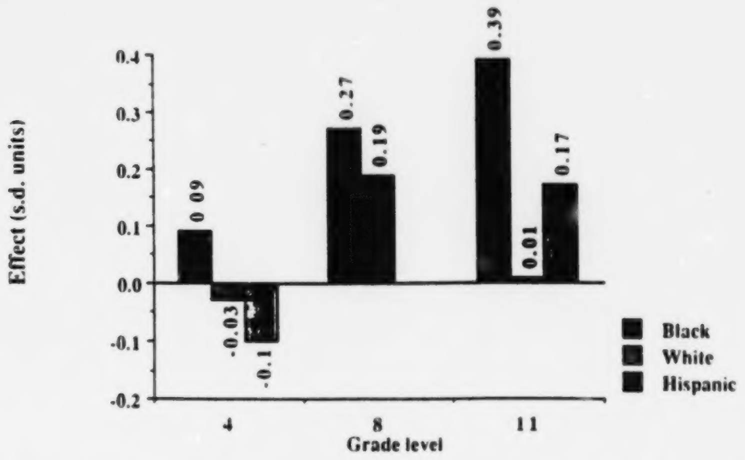

FIGURE 1. *Within-race school-level minimum competency testing program effect in standard deviation units and adjusted for control variables.*

White students, accounted for a negligible portion of the effect for Hispanic students, and caused the effect for Black students to become larger. This increase was a case of statistical suppression. A specific remedial program in the school, although not correlated with proficiency ($r = .01$), was correlated with having an MCT program ($r = .70$) and thus adds irrelevant variance to the variable MCT and reduces the relationship with proficiency. Instructional dollars per pupil, although not correlated with proficiency ($r = .04$), was correlated with MCT ($r = .27$) and acted in a similar manner. The statistical explanation is that the inclusion of these two variables in the equation suppresses the unwanted variance in reading proficiency and increases the relationship between proficiency and MCT. (For a discussion of suppression in complex regression models, see Cohen & Cohen, 1983.) Theoretically, this finding is troublesome because school-level MCT programs are accompanied by extra instructional resources and a remedial program. We know little concerning implementation effects or student achievement outcomes attributable to these factors.

Discussion

In general, these results suggested a positive relationship between school-level MCT programs and reading proficiency at the upper grade levels, but not at the elementary school level. The discussion of why this is the case is quite speculative because of the nature of the study.[2]

Grade 4

The failure to find a relationship at fourth grade for any of the race/ethnic groups suggests that there may be little or no advantage in implementing MCT programs at this grade level. In elementary schools, there is a general emphasis on instruction in basic skills, particularly reading, so the addition of an MCT program may be redundant. Because only one time point is being examined, the direction of causality between the variables cannot be established. Schools with MCT had students with lower reading proficiency, and perhaps this situation resulted in schools' implementing a local MCT program. Additionally, many other important variables (e.g., classroom practices, academic engaged time, and content covered) are critical in explaining students' reading proficiency but were not included in these analyses (Winfield, 1987b). It appears that policymakers who advocate testing reforms at this level might better use resources to strengthen teaching and instruction.

Grade 8

At eighth grade, a positive relationship between school-level MCT programs and reading was isolated for White and Black

APPENDIX

NAEP 1983–1984 Regression equation for each grade group by race

	White						Black						Hispanic					
	Grade 4		Grade 8		Grade 11		Grade 4		Grade 8		Grade 11		Grade 4		Grade 8		Grade 11	
Variable	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*
Sex[a]	3.40	0.81	6.75	0.71	6.20	1.49	7.85	1.71	6.22	1.48	6.20	1.48	4.33	2.15	5.81	1.84	1.72	1.63
Student age	−11.51	0.76	−6.92	0.82	−9.73	0.95	−7.82	1.36	−7.31	0.98	−9.73	0.96	−8.94	1.81	6.27	1.42	14.50	1.07
Region																		
Central	−1.46	1.23	1.31	1.30	3.15	2.79	6.43	2.92	8.75	2.43	3.15	2.56	−3.16	3.88	−10.43	4.23	2.79	5.73
Southeast	−2.29	1.43	2.52	1.14	−3.07	2.36	2.93	2.98	6.65	12.81	−3.07	2.36	3.84	3.30	7.15	2.97	1.82	2.73
Northeast[b]	−1.80	1.21	1.42	1.25	7.10	2.75	0.20	3.05	11.02	2.56	7.10	2.70	9.32	3.72	8.92	4.25	2.02	2.79
Student academic behaviors	1.44	0.25	1.29	0.23	1.26	0.39	0.13	0.45	1.53	0.44	1.26	0.39	0.52	0.65	0.97	0.54	2.34	0.41
Family background	3.56	0.20	5.68	0.23	4.02	0.48	1.77	0.39	3.04	0.38	4.02	0.48	3.04	0.51	3.31	0.49	4.52	0.49
Percent students free lunch	−0.05	0.01	0.06	0.01	0.03	0.04	−0.06	0.03	−0.16	−0.04	0.03	0.05	−0.12	0.04	−0.14	0.04	0.10	0.05
Percent White students	0.13	0.02	0.08	0.02	0.07	0.04	0.14	0.03	0.03	0.03	−0.01	0.06	0.16	0.05	0.03	0.05	0.09	0.04
School-level MCT	−1.08	1.08	7.79	5.54	0.42	1.58	2.89	3.97	7.60	3.23	12.34	3.48	3.00	3.23	0.06	3.71	5.76	4.33
Instructional dollars/pupil	−0.03	0.05	0.13	0.02	−0.17	0.06	0.00	0.04	0.18	0.07	−0.27	0.08	0.17	0.05	−0.13	0.07	0.13	0.05
Remedial program[c]	−1.56	0.89	4.27	0.85	6.31	2.81	−4.62	2.87	5.18	1.95	6.31	2.81	6.29	2.42	2.01	2.48	2.60	3.28

Note. NAEP = National Assessment of Educational Progress.

[a] 1 = female; 0 = male. [b] West coded "Other." [c] 1 = yes; 0 = no.

students, but not Hispanic students. One possible explanation for finding effects on Black and White students is that by restricting the variance in background factors to within groups, the probability increased of demonstrating an effect due to school variables. This should also hold for fourth grade, but unreliability of student self-report measures may be a source of error. Other studies, using a similar method, have found that among schools attended by students of the same SES background or race, it was possible to identify some that were consistently "effective or ineffective" (Frederiksen, 1975).

An alternative explanation is that there may be other factors associated with the program but not measured (e.g., monitoring of student progress) which contribute to this effect. The distributions of White and Black students' NAEP proficiency levels in schools with MCT programs were shifted upward for all students, not just for those students at the lower reading proficiency levels (Winfield, 1987a). Additional descriptive analyses of the sample schools indicated a higher percentage of students in both gifted and remedial reading programs compared to schools without MCT programs (Winfield, 1988). This suggests that the identified "MCT effect" may not be due solely to MCT but to other school-related conditions and characteristics.

The failure to find an MCT effect on the reading proficiency of Hispanic eighth graders suggests that the variables included in the analysis may be insufficient to explain reading proficiency of Hispanic students at this grade level. Other variables might influence proficiency for these students—for example, language dominance, language spoken in the home and in peer groups, and years of residence in the United States (Ortiz, 1986). Alternatively, the school variables included may operate differently in different contexts. Hanushek (1970) found differences in teachers and classrooms related to the achievement of White students but not Mexican students. These results suggest the need to investigate effects of school-level MCT for Hispanic students at this grade level.

Grade 11

At 11th grade, the MCT school effect on White students' reading proficiency could be statistically explained by the inclusion of school-level remedial program and instructional dollars per pupil. For Blacks and Hispanics, this was not the case. Remedial program and instructional dollars were suppressor variables in the regression equation for Blacks. Other research suggests that remedial programs for MCT may be less effective in facilitating reading achievement for Black students than for White students (Serow, 1984). However, remedial program in this study was a school-level rather than a student-level variable, and therefore it cannot adequately address this issue. The results of this study are merely suggestive that MCT remedial programs may have different effects on different groups.

The positive effects isolated at 11th grade for each race/ethnic group may be due to proximity to graduation. MCT may be more meaningful to students at this grade level, especially if it must be passed in order to graduate. Alternatively, the effects may be due to other unmeasured characteristics of these particular schools. Information on school retention or dropout rates was not available, and it may be that schools that have institutionalized an MCT program have higher levels of dropout among lower performing students (Serow & Davies, 1982). Thus, those students who are doing poorly or have failed an MCT may no longer be in school at 11th grade. In general, Black and Hispanic dropout rates are higher than those of White students (Plisko & Stern, 1985), so the results obtained for the 11th grade may pertain to a more select population than the populations in the 4th and 8th grades. This may be especially true of minority groups. Burton and Jones (1982) reached a similar conclusion regarding NAEP data for 17-year-olds. In a comparison of achievement trends of Black and White youth, they suggested that it would not be possible to assess whether the relative improvement observed in the Black population at ages 9 and 13 persisted at age 17 because of the differential dropout rates by race and sex. An out-of-school sample of 17-year-olds is needed to assess the trend (Burton & Jones, 1982; Alexander–James, 1987). Without adequate dropout statistics for the particular sample studied, the issue of sample selectivity at 11th grade can not be resolved.

Conclusions and Implications

Because of the exploratory nature of the studies presented here, caution must be exercised in generalizing to all schools in the nation. The results suggest that relationships between MCT programs and student reading proficiency outcomes may be varied for different grades, race/ethnic groups, and types of programs. At the elementary level, no relationship was isolated for any of the race/ethnic groups studied. At 8th grade, positive effects of school-level MCT programs were isolated for White and Black students, but not Hispanic students. Similarly, at 11th grade, positive effects of school-level MCT were isolated for each race/ethnic group.

The remaining discussion will address the limitations of the study and how future research in this area might address some of these problems. One issue in conducting secondary data analysis is the appropriateness of data used for addressing questions of interest. It is generally the case that most of the available large-scale data bases have been less than optimal for conducting certain types of policy analyses (Plisko, Ginsburg, & Chaikind, 1986). The primary purpose of NAEP—to serve as "The Nation's Report Card"—has led to very careful test construction in order to measure changes in student learning and educational competence in core subject areas. Because the major goal of NAEP is to estimate population and subpopulation means, each student may answer only a few of the large number of assessment items, and an individual reading proficiency estimate is not attained. The "plausible values" used as indicators of individual reading proficiency are intermediate steps to yield consistent estimates of selected margins of the national population—specifically, gender, ethnicity, parents' education, size and type of community, age, region of the country, and grade. These are referred to as "conditioning" variables. Analyses involving any other background variables are subject to regression effects. Thus, coefficients for school variables in this study are underestimated by about 15%–20% (R. Mislevy, personal communication, August 15, 1986), and coefficients for conditioning variables are inflated. Although the substance of any conclusions derived from these studies would be essentially unchanged if all biases were removed, any effect due to a school-level variable is *extremely conservative* when using NAEP data. Refinements in the technology of NAEP since the time of the current study have minimized this problem by conditioning on a larger number of background variables, for example, as in the NAEP study of young adult literacy (Kirsch & Jungeblut, 1986). In the future, secondary bias in NAEP data may be eliminated altogether by conditioning on well-chosen linear combinations of large numbers of variables (Mislevy, 1988).

The collection of school variables has improved the value of the assessment as a policy research tool. However, for investigating relationships between MCT programs and reading proficiency, future large-scale surveys—whether NAEP or another survey—might collect additional details on the exact nature of the school-level program. At a minimum, it is important to know the purpose, special personnel or curriculum used, nature of the state mandate, the proportion of students in the schools who failed to meet requirements, and school-level retention and dropout rates. This information would permit ruling out some of the plausible rival hypotheses suggested for results reported in this study and would permit a more precise categorization of MCT programs. For example, whether a program is used for remediation funding or promotion/graduation might differentially influence student reading proficiency. The use of principals' self-reports concerning whether there is a school-level MCT program is admittedly limited in defining the nature, implementation, and practices affecting student outcomes in reading.

Any explanation of student outcomes needs to consider teachers and classroom practices, as well as school variables. The effects of these variables are subsumed somewhat in an overall school effect; however, teacher or classroom-level variables deserve direct examination. These were not within the scope of the study reported here and would be important to include in future studies. Also, information on student participation in MCT and remedial programs should be collected at the individual student level rather than at the school level. In conducting multilevel analyses of school data, it is known that many of the assumptions of using aggregated data do not reflect reality. For example, school-level aggregates are assumed to affect all students equally, but pupils receive differential exposure to school resources and facilities. This may be especially true with respect to low-achieving students who are more likely to participate in MCT-related remediation and be placed in lower-tracked classes (Braddock, 1990; Lee & Bryk, 1988). Other individual student-level data, not available in NAEP, include a measure of SES such as parents' occupational status. Future analyses should include this measure as a covariate to adjust for student background.

Large-scale, cross-sectional data as used in this study do not permit one to infer the direction of causality or to understand the nature of the process. In order to assess change in students' reading proficiency that may result from implementation of an MCT program, data must be available for two or more time periods, and preferably for cohorts of students. Moreover, school-level MCT program was used as a proxy variable for conditions within schools which influence student reading outcomes. The process variables, interactions, and local adaptations within schools which impact student outcome—such as teachers' expectations, resource allocation and use, and students' opportunity to learn—might be more appropriately studied by using qualitative methods. (See Oakes, 1989, for a discussion of educational indicators and school context.)

The diversity of local and state policies related to MCT makes it extremely difficult to characterize precisely the nature and outcomes of the testing reforms. The NAEP data were not designed to specifically address such issues; however, this study demonstrates both the limits and potential of using redesigned NAEP data to explore the relation between testing reforms and achievement from a national perspective. For policymakers interested in the use of testing to improve schools, there is modest evidence that a positive relationship exists, at least at the middle and secondary school level. We know, however, that any reform operates on the levels of policy, administration, and practice—each with its own rewards, incentives and limitations (Elmore & McLaughlin, 1988). To the extent that NAEP in its redesign can inform how teachers teach and how school organization affects practice in schools implementing MCT programs, we may begin to understand how this reform has its impact on student achievement outcomes. In the future, NAEP will conduct a state-by-state assessment of reading and mathematics achievement. The present study, with its noted limitations, foreshadows some of the problems inherent in isolating achievement effects that may result from various school reforms.

Notes

[1] The use of hierarchical liner modeling (Raudenbush & Bryk, 1988–1989) provides an optimal method for analyzing data from studies of school and classroom effects. The NAEP samples too few students within each school to provide a stable estimate.

[2] The author is aware that the implications of the study for policy are limited due to the exploratory nature of the study and use of the redesigned NAEP data. In attempting to arrive at an estimate of the relationship between testing reforms in the nation and student reading achievement by using NAEP data, one misses the local context and implementation effects—factors that are closely linked to student reading achievement outcomes. Local school- or district-level data my be more appropriate for developing policies regarding

school organization and achievement outcomes. The policy dynamics of school reform, as well as changing schools, are highly complex (Timar & Kirp, 1989). Reanalyses of NAEP data, however, may be informative in describing existing relationships among school policy, teacher behaviors, and student achievement outcomes.

References

Airasian, P. (1987). State-mandated testing and educational reform: Contexts and consequences. *American Journal of Education, 95*, 393–412.

Alexander, L., & James, T. (1987). *The nation's report card. Report of the Study Group.* Cambridge, MA: National Academy of Education.

Anderson, B., & Pipho, C. (1984). State-mandated testing and the fate of local control. *Phi Delta Kappan, 66*, 209–212.

Barr, R., & Dreeben, R. (1983). *How schools work.* Chicago: University of Chicago Press.

Braddock, J. H. (1990). Tracking implications for student race-ethnic subgroups (Rep. no. 1). Baltimore, MD: Center for Research on Effective Schooling for Disadvantaged Students.

Broudy, H. S. (1980). Impact of minimum competency testing on curriculum. In R. M. Jaeger & C. K. Tittle (Eds.), *Minimum competency achievement testing: Motives, models, measures, and consequences* (pp. 108–117). Berkeley, CA: McCutchan.

Burstein, L., & Miller, M. D. (1981). Regression-based analyses of multilevel educational data. In R. F. Boruch, P. M. Wortman, & D. S. Cordray (Eds.), *Reanalyzing program evaluations* (pp. 194–207). San Francisco, CA: Jossey-Bass.

Burton, N. W., & Jones, L. V. (1982). Recent trends in achievement levels of black and white youth. *Educational Researcher, 11*, 10–17.

Chubb, J. E. (1988). Why the current wave of school reform will fail. *The Public Interest*, No. *90*, 28–49.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Cohen, D. K., & Haney, W. (1980). Minimum competency testing and social policy. In R. M. Jaeger & C. K. Tittle (Eds.), *Minimum competency achievement testing: Motives, models, measures, and consequences* (pp. 5–22). Berkeley, CA: McCutchan.

Coleman, J. S., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., & York, R. (1966). *Equality of educational opportunity.* Washington, DC: U.S. Government Printing Office.

Congressional Budget Office. (1987). *Educational achievement: Explanations and implications of recent trends.* Washington, DC: U.S. Government Printing Office.

Darling-Hammond, L., & Wise, A. E. (1985). Beyond standardization: State standards and school improvement. *Elementary School Journal, 85*, 315–336.

Edmonds, R. R. (1979). Effective schools for the urban poor. *Educational Leadership, 37*, 15–24.

Education Commission of the States (1984). *Current status of state assessment programs.* Denver, CO: Author.

Eubanks, E. E., & Levine, D. U. (1983). *A first look at effective school projects at inner-city elementary schools.* Paper presented at the annual meeting of the American Association of School Administrators, Atlantic City, NJ.

Elmore, R. F., & McLaughlin, M. W. (1988). *Steady work, policy, practice, and the reform of American Education.* Santa Monica, CA: RAND.

Ellwein, M. C., Glass, G. V., & Smith, M. L. (1988). Standards of competence: Propositions on the nature of testing reforms. *Educational Researcher, 17*(8), 4–9.

Ellwein, M. C., Popp, R., & Niemann, M. (1988, February). *Minimum competency testing, standard-setting, rhetoric and research.* Paper presented at the 1988 Eastern Educational Research Association meeting, Miami Beach, FL.

Frederiksen, J. R. (1975). *School effectiveness and equality of educational opportunity* (Report submitted to the Carnegie Corporation of New York). Unpublished manuscript.

Fullan, M. (1985). Change processes and strategies at the local level. *Elementary School Journal, 85*, 391–421.

Glass, G. V. (1977). Integrating findings: The meta analysis of research. In L. Shulman (Ed.), *Review of research in education* (Vol. 5, pp. 351–379). Itasen, IL: F. E. Peacock.

Good, T. L., & Brophy, J. E. (1986). School effects. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 570–602). New York: MacMillan.

Goodlad, J. I. (1984). *A place called school.* New York: McGraw-Hill.

Hanushek, E. (1970). *The value of teaching.* Santa Monica, CA: RAND.

Jaeger, R. M. (1982). The final hurdle: Minimum competency achievement testing. In G. R. Austin & H. Garber (Eds.), *The rise and fall of* *national test scores* (pp. 223–246). New York: Academic Press.

Jencks, C. S., Smith, M., Acland, H., Bane, M. J., Cohen, D., Gintis, H., Heyns, B., & Michelson, S. (1972). *Inequality: A reassessment of the effect of family and schooling in America.* New York: Basic Books.

Johnson, E. G. (1987). Design effects. In A. E. Beaton (Ed.), *Implementing the new design: The NAEP 1983–84 technical report* (Report No. 15-TR-20). Princeton, NJ: Educational Testing Service.

Jonas, E., & Wallace, L. (1986, April). *Minimum competency testing: An analysis of student outcomes for those not mastering mandated testing standards.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Kirsch, I. S., & Jungeblut, A. (1986). *Literacy: Profiles of America's young adults* (Final Report No. 16-PL-01). Princeton, NJ: National Assessment of Educational Progress.

Kish, L., & Frankel, M. R. (1974). Inference from complex samples. *Journal of the Royal Statistical Society,* Series B, *36,* 1–22.

Koretz, D. (1988). Arriving in Lake Wobegone: Are standardized tests exaggerating achievement and distorting instruction? *American Educator, 12*(2), 8–15, 46–52.

Kyle, R. M. J. (Ed.). (1985). *Reaching for excellence: An effective schools sourcebook.* Washington, DC: U.S. Government Printing Office.

Lee, V., & Bryk, A. S. (1988). Curriculum tracking as mediating the social distribution of achievement in Catholic and public secondary schools. *Sociology of Education,* 61, 78–94.

Linn, R. L., Madaus, G. F., & Pedulla, J. (1982). Minimum competency testing: Cautions on the state of the art. *American Journal of Education, 91,* 15–19.

MacKenzie, D. E. (1983). Research for school improvement: an appraisal of some recent trends. *Educational Researcher, 12*(4), 5–17.

Madaus, G. F., Airasian, P. W., & Kellaghan, T. (1980). *School effectiveness: A reassessment of the evidence of school effectiveness.* New York: McGraw-Hill.

Mangino, E., & Babcock, M. A. (1986, April). *Minimum competency testing: Helpful or harmful to high-level skills.* Paper presented at the Annual Meeting of the American Educational Research Association Meeting, San Francisco.

Mayeske, G. W., Wisler, C. E., Beaton, A. E., Weinfeld, F. D., Cohen, W. M., Okada, T., Proshek, J. M., & Tabler, K. A. (1972). *A study of our nation's schools* (DHEW Publication No. [OE] 72–142). Washington, DC: U.S. Department of Health, Education, and Welfare.

Mayeske, G. W., Okada, T., Cohen, W. M., Beaton, A., & Wisler, C. E. (1973). *A study of the achievement of our nation' students* (DHEW Publication No. [OE] 72–131). Washington, DC: U.S. Department of Health, Education, and Welfare.

Messick, S., Beaton, A., & Lord, F. (1983). *National Assessment of Educational Progress redesigned: A new design for a new era* (NAEP Report 83–1). Princeton, NJ: National Assessment of Educational Progress.

Mislevy, R. (1985). *Inferences about the latent populations from complex samples* RR 85–41). Princeton, NJ: Educational Testing Service.

Mislevy, R. J. (1988). *Randomization-based inferences about latent variables from complex samples.* Princeton, NJ: Educational Testing Service.

Mosteller, F., & Moynihan, D. P. (1972). A pathbreaking report. In F. Mosteller and D. P. Moynihan (Eds.), *On equality of educational opportunity.* New York: Vintage Books.

Murphy, J. (1989). Educational reform in the 1980's: Explaining some surprising success. *Educational Evaluation and Policy Analysis, 11,* 209–221.

National Assessment of Educational Progress. (1985). *The reading report card.* Princeton, NJ: Educational Testing Service.

National Assessment of Educational Progress (1986a). *National Assessment of Educational Progress 1982–84: A technical report.* Princeton, NJ: Educational Testing Service.

National Assessment of Educational Progress (1986b). *User's Guide 1983–84 public-use data tapes, Version 3.1.* Princeton, NJ: Educational Testing Service.

Ortiz, V. (1986). Reading activities and reading proficiency among Hispanic, Black, and White students. *American Journal of Education,* 95(1) 58–76.

Pipho, C. (1983). *Student minimum competency testing* (Issuegram 20.) Denver, CO: Education Commission of the States. (ERIC Document Reproduction Service No. ED 234 091).

Pipho, C., & Hadley, C. (1984). Clearinghouse notes, state activity: Minimum competency testing as of December 1984. Denver, CO: Education Commission of the States.

Plisko, V., Ginsburg, A., & Chaikind, S. (1986). Assessing national data on education. *Educational Evaluation and Policy Analysis, 8,* 1–16.

Plisko, V., & Stern, J. (1985). *The condition of* *education.* Washington, DC: U.S. Government Printing Office.

Purkey, S. C., & Smith, M. S. (1983). Effective schools: A review. *The Elementary School Journal, 83*(4), 427–452.

Purkey, S. C., & Smith, M. S. (1985). School reform: The district policy implications of the effective schools literature. *The Elementary School Journal, 85*, 353–389.

Raudenbush, S., & Bryk, A. S. (1988–1989). Methodological advances in analyzing the effects of schools and classrooms on student learning. In E. Z. Rothkopf (Ed.), *Review of research in education, 15* (chap. 10). Washington, DC: American Educational Research Association.

Serow, R. (1984). Effects of minimum competency testing for minority students: A review of expectations and outcomes. *The Urban Review, 16*(2), 67–75.

Serow, R., & Davies, J. (1982). Resources and outcomes of minimum competency testing as measures of equality of educational opportunity. *American Educational Research Journal, 19*, 529–539.

Sirotnik, K. A., & Clark, R. W. (1988). School-centered decision making and renewal. *Phi Delta Kappan, 69*, 660–664.

Sirotnik, K. A., & Oakes, J. (1986). Critical inquiry for school renewal: Liberating theory and practice. In K. A. Sirotnik & J. Oakes (Eds.), *Clinical perspectives on the organization and improvement of schooling* (chap. 1). Boston: Kluwer-Nijhoff.

Smith, M.S. (1972). Equality of educational opportunity: The basic finding reconsidered. In F. Mosteller & D. P. Moynihan (Eds.), *On equality of educational opportunity*. New York: Vintage Books.

Stedman, L. C. (1987). It's time we changed the effective schools formula. *Phi Delta Kappan, 69*, 215–224.

Stringfield, S., & Teddlie, C. (1988). A time to summarize: The Louisiana school effectiveness study. *Educational Leadership,, 46*, 43–49.

Timar, T., & Kirp, D. (1989). Education reform in the 1980's: Lessons from the states. *Phi Delta Kappan, 70*, 504–511.

Venezky, R. L., & Winfield, L. F. (1979). *Schools that succeed beyond expectations in reading* (Studies on Education, Tech. Rep. No. 1). Newark: University of Delaware. (ERIC Document Reproduction Service No. ED 177 484)

Winfield, L. F. (1987a). *The relationship between minimum competency testing programs and student reading proficiency: Implications from the 1983–1984 National Assessment of Educational Progress in Reading and Writing* (ETS Research Report No. 87-115). Princeton, NJ: Educational Testing Service.

Winfield, L. F. (1987b). Teachers' estimates of test content covered in class and first grade students' reading achievement. *Elementary School Journal, 87*, 437–454.

Winfield, L. F. (1988, April). *Characteristics of schools implementing minimum competency testing programs.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Author

LINDA F. WINFIELD, Principal Research Scientist, Center for Research on Effective Schooling for Disadvantaged students, The Johns Hopkins University, 3505 N. Charles St., Baltimore, MD 21218. *Specializations*: social organization of schools: school characteristics and policies that impact reading/literacy attainment.

School Practices

Review of Educational Research
Fall 1990, Vol. 60, No. 3, pp. 471–499

Achievement Effects of Ability Grouping in Secondary Schools: A Best-Evidence Synthesis

Robert E. Slavin
Johns Hopkins University

This article reviews research on the effects of ability grouping on the achievement of secondary students. Six randomized experiments, 9 matched experiments, and 14 correlational studies compared ability grouping to heterogeneous plans over periods of from one semester to 5 years. Overall achievement effects were found to be essentially zero at all grade levels, although there is much more evidence regarding Grades 7–9 than 10–12. Results were similar for all subjects except social studies, for which there was a trend favoring heterogeneous placement. Results were close to zero for students of all levels of prior performance. This finding contrasts with those of studies comparing the achievement of students in different tracks, which generally find positive effects of ability grouping for high achievers and negative effects for low achievers, and these contrasting findings are reconciled.

For more than 70 years, ability grouping has been one of the most controversial issues in education. Its effects, particularly on student achievement, have been extensively studied over that time period, and many reviews of the literature have been written. In recent years, a comprehensive review of the achievement effects of ability grouping in elementary schools has been published by Slavin (1987), but only brief meta-analyses by Kulik and Kulik (1982, 1987) have reviewed the evidence on ability grouping and heterogeneous placement in secondary schools.

The purpose of this paper is to present a comprehensive review of all research published in English that has evaluated the effects of ability grouping on student achievement in secondary schools. *Secondary schools* are defined here as middle, junior, or senior high schools in the United States, or similarly configured secondary schools in other countries. Secondary schools can include grades as low as five, but they usually begin with sixth or seventh grades. Ability grouping is defined as any school or classroom organization plan that is intended to reduce the heterogeneity of instructional groups; in between-class ability grouping the heterogeneity of each class for a given subject is reduced, and in within-class ability grouping the heterogeneity of groups within the class (e.g., reading groups) is reduced.

Unlike the situation in elementary schools, the type of ability grouping used in

---

This paper was written under funding from the Office of Educational Research and Improvement, U.S. Department of Education, as part of a subcontract with the National Center on Effective Secondary Schools at the University of Wisconsin. However, any opinions expressed are those of the author and do not represent OERI positions or policy. I would like to thank Fred Newmann, Jeannie Oakes, Adam Gamoran, Harry Passow, and Alan Kerckhoff for their comments on an earlier draft and Paula Thomas for her assistance in locating references.

Copyright 1990 by the American Educational Research Association. Reprinted by permission of the publisher.

secondary schools is overwhelmingly between-class grouping (McPartland, Coldiron, & Braddock, 1987). Several closely related forms of ability grouping are used. Sometimes students are assigned on the basis of some combination of composite achievement, IQ, and teacher judgments to a track, within which all courses are taken. For example, senior high school students are often assigned to academic, general, and vocational tracks; middle/junior high school students are often assigned to advanced, basic, and remedial tracks (in either case, the number of tracks and the names used to describe them vary widely). This type of grouping plan is generally called tracking in the United States or streaming in Europe. It is an example of what Slavin (1987) called "ability-grouped class assignment." In addition to assignment to higher and lower sections of the same courses, tracking in senior high schools usually involves different courses or course requirements. For example, a student in the academic track may be required to take more years of mathematics than a student in the general track, or may take French III rather than metal shop.

A particular form of tracking often seen in middle/junior high schools is block scheduling, where students spend all or most of the day with one homogeneous group of students. Some schools rank-order students from top to bottom and assign them to, say, 7-1, 7-2, 7-3, and so on. Many senior high schools allow students to choose their track or to choose the level they wish to take in each subject, but in plans of this kind counselors tend to steer students into the level of classes to which they would have been assigned if the school were not allowing students a choice (Rosenbaum, 1978).

Another form of ability grouping common in secondary schools involves assigning students to ability-grouped classes for all academic subjects, but allowing for the possibility that students will be placed in a high-ranking group for one subject and a low-ranking group for another. In practice, scheduling constraints often make this type of grouping similar to plans in which all courses are taken within the same track. In some cases schools ability group for some subjects and not for others; for example, students may be in ability-grouped math and English classes but in heterogeneous social studies and science classes. Ability grouping usually involves higher and lower sections of the same course, but sometimes consists of assignment to completely different courses, as when ninth graders are assigned either to Algebra I or to general math. When high achievers are assigned to markedly different courses usually offered to older students (as when seventh graders take algebra), this is called acceleration. More commonly, high achievers may be assigned to "honors" or "advanced placement" sections of a given course, and low achievers may be assigned to special "remedial" sections.

Although between-class ability grouping is by far the most common type of ability grouping in secondary schools, forms of within-class grouping are also occasionally seen. These are plans in which students are assigned to homogeneous instructional groups within their classes. Within-class ability grouping, such as use of reading or math groups, is the most common form of grouping at the elementary level (McPartland et al., 1987). Complex plans, such as those that involve grouping across grade lines, flexible grouping for particular topics, and part-time grouping, are also occasionally seen in secondary schools. In general, a wider range of grouping plans are used in middle/junior high schools than in senior high schools.

Arguments for and against ability grouping have been essentially similar for 70 years. For example, Turney (1931), summarizing writings of the 1920s, listed advantages and disadvantages of ability grouping. The advantages were as follows:

1. It permits pupils to make progress commensurate with their abilities.
2. It makes possible an adaption of the technique of instruction to the needs of the group.
3. It reduces failures.
4. It helps to maintain interest and incentive, because bright students are not bored by the participation of the dull.
5. Slower pupils participate more when not eclipsed by those much brighter.
6. It makes teaching easier.
7. It makes possible individual instruction to small slow groups.

The following were the disadvantages:

1. Slow pupils need the presence of the able students to stimulate them and encourage them.
2. A stigma is attached to low sections, operating to discourage the pupils in these sections.
3. Teachers are unable, or do not have time, to differentiate the work for different levels of ability.
4. Teachers object to the slower groups.

A research symposium, school board meeting, or PTA meeting on the topic of ability grouping in 1990 is likely to bring up much the same arguments on both sides, with two important additions: the argument that ability grouping discriminates against minority and lower-class students (e.g., Braddock, 1990; Rosenbaum, 1976), and the argument that students in the low tracks receive a lower pace and lower quality of instruction than do students in the higher tracks (e.g., Gamoran, 1989; Oakes, 1985).

In essence, the argument in favor of ability grouping is that it will allow teachers to adapt instruction to the needs of a diverse student body and give them an opportunity to provide more difficult material to high achievers and more support to low achievers. The challenge and stimulation of other high achievers are believed to be beneficial to high achievers (see Feldhusen, 1989). Arguments opposed to ability grouping focus primarily on the perceived damage to low achievers, who receive a slower pace and lower quality of instruction, have teachers who are less experienced or able and who do not want to teach low-track classes, face low expectations for performance, and have few positive behavioral models (e.g., Gamoran, 1989; Oakes, 1985; Persell, 1977; Rosenbaum, 1980). Because of the demoralization, low expectations, and poor behavioral models, students in the low tracks are believed to be more prone to delinquency, absenteeism, dropout, and other social problems (Crespo & Michelna, 1981; Wiatrowski, Hansell, Massey, & Wilson, 1982). With few college-bound peers, students in low tracks have been found to be less likely to attend college than other students (Gamoran, 1987). Ability grouping is perceived to perpetuate social class and racial inequities because lower class and minority students are disproportionally represented in the lower tracks. Ability grouping is often considered to be a major factor in the development of elite and under-class groups in society (Persell, 1977; Rosenbaum, 1980). Perhaps most important, tracking is believed to work against egalitarian, democratic ideals by sorting students into categories from which escape is difficult or impossible.

There are important differences between the pro-grouping and anti-grouping positions that go beyond the arguments themselves. Arguments in favor of ability grouping focus on *effectiveness*, saying in effect that as distasteful as grouping may be, it so enhances the learning of students (particularly but not only high achievers) that its use is necessary. In contrast, arguments opposed to grouping focus at least as much on *equity* as on effectiveness and on democratic values as much as on outcomes. In one sense, then, the burden of proof is on those who favor grouping, for if grouping is not found to be clearly more effective than heterogeneous placement, none of the pro-grouping arguments apply. The same is not true of anti-grouping arguments, which provide a rationale for abolishing grouping that would be plausible even if grouping were found to have no adverse effect on achievement.

Research on the achievement effects of ability grouping has taken two broad forms. One type of research compares the achievement gains of students who are in one or another form of grouping to those of students in ungrouped, heterogeneous placements. Another type of research compares the achievement gains made by students in high-ability groups to those made by students in the low groups.

Reviews of the grouping versus nongrouping literature have consistently shown that grouping has little or no impact on overall student achievement in elementary and secondary schools (e.g., Borg, 1965; Esposito, 1973; Findley & Bryan, 1971; Good & Marshall, 1984; Heathers, 1969; Kulik & Kulik 1987). Primarily on the basis of his own empirical research, Borg (1965) claimed that ability grouping had a slight positive effect on the achievement of high achievers and a slight negative effect on low achievers, but Kulik and Kulik (1987) found no such trend.

In contrast, researchers who have compared gains made by students in different tracks have generally concluded that controlling for ability level, socioeconomic status, and other control variables, being in the top track accelerates achievement and being in the low track significantly reduces achievement (Alexander, Cook, & McDill, 1978; Dar & Resh, 1986; Gamoran & Berends, 1987; Gamoran & Mare, 1989; Oakes, 1982; Persell, 1977; Sorensen & Hallinan, 1986). In fact, many researchers and theorists in the sociological tradition maintain that tracking is a principal source of social inequality in society and that it causes or greatly magnifies differences along lines of class and ethnicity (e.g., Braddock, 1990; Jones, Erickson, & Crowell, 1972; Schafer & Olexa, 1971; Vanfossen, Jones, & Spade, 1987).

One area of research has investigated the quality of instruction offered to students in high- and low-ability groups, usually concluding that low-ability group classes receive instruction that is significantly lower in quality than that received by students in high-track classes (e.g., Evertson, 1982; Gamoran, 1989; Oakes, 1985; Trimble & Sinclair, 1987). However, it is difficult to compare "quality of instruction" in high- and low-track classes. For example, teachers typically cover less material in low-track classes (e.g., Oakes, 1985). Is this an indication of poor quality of instruction or an appropriate pace of instruction? Students in low-track classes are more off-task than those in high-track classes (e.g., Evertson, 1982). Is this due to the poor behavioral models and low expectations in the low-track classes, or would low achievers be more off-task than high achievers in any grouping arrangement? Evidence that low-track classes are often taught by less experienced or less qualified teachers or that they manifest other objective indicators of lower-quality instruction could justify the conclusion that regardless of measurable effects on learning, students in the lower tracks do not receive equal treatment, but such evidence is rare.

In addition to synthesizing research on overall effects of ability grouping on the achievement of high-average- and low-achieving secondary students, this review will attempt to reconcile research comparing achievement gains in different tracks with research comparing grouped and ungrouped settings.

Review Methods

This review uses a procedure called "best-evidence synthesis" (Slavin, 1986), which incorporates the best features of meta-analytic and traditional reviews. Best-evidence syntheses specify clear, well-justified methological and substantive criteria for inclusion of studies in the main review and describe individual studies and critical research issues in the depth typical of good-quality narrative reviews. However, whenever possible, effect sizes are used to characterize study outcomes, as in meta-analyses (Glass, McGaw, & Smith, 1981). Systematic literature search procedures, also characteristic of meta-analysis, are similarly applied in best-evidence syntheses.

Criteria for Study Inclusion

The studies on which this review is based had to meet a set of a priori criteria with respect to relevance to the topic and methodological adequacy. First, all studies had to involve comprehensive ability grouping plans that incorporated most or all students in the school. This excludes studies of special programs for the gifted or other high achievers as well as studies of special education, remedial programs, or other special programs for low achievers. Studies of within-class ability grouping are included, but studies of such grouping-related programs as individualized instruction, mastery learning, cooperative learning, and continuous-progress groupings are excluded.

Studies had to be available in English, but otherwise no restrictions were placed on study location or year of publication. Every attempt was made to locate dissertations and other unpublished documents in addition to the published literature.

Methodological requirements for inclusion. Criteria for inclusion of studies in the main review were essentially identical to those used in an earlier review of elementary ability grouping (Slavin, 1987). These were as follows:

1. Ability-grouped classes were compared to heterogeneously grouped classes. This requirement excluded a few studies that correlated "degree of heterogeneity" with achievement gain (e.g., Millman & Johnson, 1964; Wilcox, 1963). Studies that compared achievement gains for students in different tracks but not to heterogeneous classes (e.g., Alexander et al., 1978) were excluded from the main review but are discussed in a separate section.

2. Achievement data from standardized or teacher-made tests were presented. This excluded many anecdotal reports and studies that used grades as the dependent measure. Teacher-made tests, used in a very small number of studies, were accepted

only if there was evidence that they were designed to assess objectives taught in all classes.

3. Initial comparability of samples was established by use of random assignment or matching of students or classes. When individual students in intact schools or classes were matched, evidence had to be presented that the intact groups were comparable.

4. Ability grouping had to be in place for at least a semester.

5. At least three ability-grouped and three control classes were involved.

The criteria outlined above excluded very few studies comparing comprehensive ability grouping plans to heterogeneous placements. Every study located that satisfied criteria 1, 2, and 3 also satisfied criteria 4 and 5. Excluding studies of special programs for high achievers (e.g., Atkinson & O'Connor, 1963), all but two of the studies included in meta-analyses by Kulik and Kulik (1982, 1987) were also included in the present review. The exceptions were a study by Adamson (1971) that had substantial IQ differences favoring the ability-grouped school and one by Wilcox (1963) that compared more and less heterogeneous tracked classes.

One major category of studies included in the present review but excluded by the Kuliks includes studies that did not present data from which effect scores could be computed (e.g., Borg, 1965; Ferri, 1971; Lovell, 1960; Postlethwaite & Denton, 1978). These studies are discussed in terms of the direction and statistical significance of their findings.

Literature Search Procedures

The studies included here were located in an extensive search. Principal sources included the Education Resources Information Center (ERIC), *Dissertation Abstracts International*, and citations made in other reviews, meta-analyses, and primary sources. Every attempt was made to obtain a complete set of published and unpublished studies that met the criteria outlined above.

Computation of Effect Sizes

Effect sizes were generally computed as the difference between the experimental and control means divided by the control group's standard deviation (Glass et al., 1981). In the ability grouping literature, the heterogeneous group is almost always considered the control group, and this convention is followed in the present article; positive effect sizes are ones that favored ability grouping, whereas negative effect sizes indicated higher means in the heterogeneous groups. The standard deviation of the heterogeneous group is also preferred as the denominator because of the possibility that ability grouping may alter the distribution of scores. However, when means or standard deviations were omitted in studies that otherwise met the inclusion criteria, effect sizes were estimated when possible from *t*s, *F*s, exact *p* values, sums of squares in factorial designs, or other information, following procedures described by Glass et al. (1981).

Several of the studies included in this review presented data comparing gain scores without reporting actual pre- or posttest means. Standard deviations of gain scores are typically lower than those of raw scores (to the degree that pre-post correlations exceed +0.5), so effect sizes computed on gain scores are often inflated. If pre-post correlations are known, effect sizes from all scores can be transformed to the scale of posttest values. However, because none of the studies using gain scores also provided pre-post correlations, a pre-post correlation of +0.8 was assumed (following Slavin, 1987). Using a formula from Glass et al. (1981), this correlation produces a multiplier of 0.632, which was used to deflate effect size estimates from gain score data. The purpose of this procedure and others was to attempt to put all effect size estimates in the same metric, the unadjusted standard deviation of the heterogeneous classes. However, because this multiplier is only a rough approximation, effect sizes from studies using gain scores should be interpreted with even more caution than that which is warranted for effect sizes in general.

Another deviation from usual meta-analytic procedure used in the present view involved adjustments of posttest scores for any pretest differences. This was done either by subtracting pretest means from posttests (if the same tests were used), by converting pre- and posttest means to *z* scores and then subtracting (if different tests were used), or by using covariance-adjusted scores. However, even when such adjustments were made, affecting the numerator of the effect size formula, the denominator remained the unadjusted posttest standard deviation.

One effect size is reported for each study (see Bangert-Drowns, 1986). When multiple subsamples, subjects, or tests were used, medians were computed across the data points. For example, if four measures were used with three subgroups (e.g., high, middle, and low achievers), the effect size for the study as a whole would be the median of the 12 (4 × 3) resulting effect sizes. Whenever possible, findings were also broken down by achievement level (high, average, low), and separate effect sizes were computed for each major subject.

In pooling findings across studies, medians rather than means were used, principally to avoid giving too much weight to outliers. However, any measure of central tendency in a meta-analysis or best-evidence synthesis should be interpreted in light of the quality and consistency of the studies from which it was derived, not as a finding in its own right.

Research on Ability Grouping in Secondary Schools

A total of 29 studies of tracking or streaming in secondary schools met the inclusion criteria listed earlier. The studies, their major characteristics, and their findings are listed in Table 1.

The studies listed in Table 1 are organized in three categories according to their research designs. Six studies used random assignment of students to ability-grouped or heterogeneous classes. Nine studies took groups of students; matched them individually on IQ, composite achievement, and other measures; and then assigned one of each matched pair of students to an ability-grouped class and one to a heterogeneous class. The quality of these randomized or matched experimental designs is very high, and the findings of the 15 studies using such designs must be given special weight. The remaining 14 studies investigated existing schools or classrooms that used or did not use ability grouping, and then either selected matched groups of students from within each type of school or used analyses of covariance or other statistical procedures to equate the groups. The difficulty inherent in such designs is that any differences between schools that are systematically related to ability grouping would be confounded with the practice of ability

TABLE 1
Studies of secondary tracking

Article	Grades	Location	Sample size	Duration	Design	Effect Sizes		
						By achievement	By subject	Total
					Randomized experimental studies			
Marascuilo & McSweeney, 1972	8–9	Berkeley, CA	603 students	2 yrs.	Students randomly assigned to 3-group AG or hetero social studies classes. Compared students on teacher-made and standardized tests.	HI +.14 AV −.37 LO −.43	Social studies −.22	−.22
Drews, 1963	9	Lansing, MI	4 schools, 432 students	1 yr.	Students randomly assigned to 3-group AG or hetero English classes compared on standardized tests.	HI −.16 AV +.01 LO −.01	Reading −.11 Language .00	−.05
Fick, 1963	7	Olathe, KS	1 school, 168 students	1 yr.	Students randomly assigned to 3-group AG or hetero "core" classes. Both classes taught by same teacher. Iowa Tests of Basic Skills used as posttests.	HI +.01 AV .00 LO −.04	Reading −.01 Language .00	−.01
Peterson, 1966	7–8	Chisholm, MN	1 school, 152 students	1 yr.	Students randomly assigned to AG or hetero classes. AG based on composite achievement, grades. Compared on standardized tests.	HI +.05 AV −.44 LO −.06	Reading +.02 Language −.01 Math −.25 Social studies −.07	−.04

Ford, 1974	9	New York, NY (Low SES minority)	80 students	1 sem.	Students randomly assigned to 2-group AG or hetero math classes. Same teachers taught both types of classes. Students compared on Metropolitan Achievement Test.		Math (0)	(0)
Bicak, 1962	8	Minneapolis, MN (lab school)	1 school, 75 students	1 sem.	Students randomly assigned to 2-group AG or hetero science classes at university lab school.	HI −.39 LO −.10	Science −.25	−.25
Matched experimental studies								
Lovell, 1960	10	Panama City, FL	500 students	1 yr.	Matched students assigned to 5-group AG or hetero English, biology, and algebra classes	HI (+) AV (0) LO (0)	Language (+) Math (0) Biology (0)	(0)
Billett, 1928	9	Painesville, OH	408 students	1 yr.	In 3 successive years, matched students assigned to 3-group AG or hetero English classes. Compared gains on standardized tests.	HI −.11 AV +.03 LO +.18	English +.04	+.04
Platz, 1965	9	—[a]	298 students	1 sem.	Matched students assigned to 3-group AG or hetero science classes. Students compared on standardized science test.	HI +.24 AV −.10 LO +.22	Science +.22	+.22
Bailey, 1968	9	St. Louis, MO	255 students	1 yr.	Matched students assigned to 2-group AG or hetero algebra classes. Same teachers taught both types of classes. Students compared on gains on standardized algebra measure.	HI +.18 LO −.24	Math −.03	−.03

continued on p. ***III - 84***

TABLE 1 (continued)

Article	Grades	Location	Sample size	Duration	Design	Effect Sizes: By achievement	Effect Sizes: By subject	Effect Sizes: Total
Thompson, 1974	11	Suburban VA	240 students	1 yr.	Compared students in 2 schools, one AG in social studies, one hetero. Students matched. Compared gain scores on teacher-made tests.	HI −.50 HI AV −.47 LO AV −.43 LO −.54	Social studies −.48	−.48
Barton, 1964	9	Rural UT	204 students	1 yr.	Matched students assigned to 4-group AG or hetero English classes, compared on California Achievement Test gains.	HI +.22 HI AV −.03 LO AV −.13 LO −.20	Reading +.06 Language −.13	−.04
Willcutt, 1969	7	Bloomington, IN (lab school)	156 students	1 yr.	Matched students assigned to 4-group flexible AG in math or to hetero. Grouping changed 8 times in the year.		Math −.15	−.15
Holy & Sutton, 1930	9	Marion, OH	148 students	1 sem.	Matched students assigned to AG, hetero algebra classes. Same teacher taught all classes.		Math +.28	+.28
Martin, 1927	7	New Haven, CT	83 students	1 yr.	Matched students assigned to 3-group AG or hetero.	HI +.12 AV −.06 LO +.23	Reading +.17 Math +.13 Language +.03	+.10

							Social studies .00 Science −.04	
				Correlational studies				
Kerckhoff, 1986	5–10	Britain	8,500 students	5 yrs.	Longitudinal study of students throughout Britain who attended streamed or unstreamed secondary schools.		Reading +.02 Math +.03	+.03
Fogelman, Essen, & Tibbenham, 1978	6–10	Britain	5,923 students	4 yrs.	Retrospective study compared students who had been in streamed, partially streamed, or heterogeneous schools throughout secondary school, controlling for Grade 5 general ability.		Reading +.02 Math +.03	+.03
Borg, 1965	6–9 7–10 8–11 9–12	Utah	2,934 students	4 yrs.	Longitudinal study of students in districts using AG compared to students in neighboring district using heterogeneous grouping, controlling for pretests.	HI (0) AV (0) LO (0)	Math (0) Science (0)	(0)
Ferri, 1971	5–6	Britain	28 schools, 1,716 students	2 yrs.	Streamed and nonstreamed schools matched on 7+ (Grade 2) reading, followed 4 years in junior school, 2 years in secondary.	HI (0) AV (0) LO (0)	Math (0) English (0)	(0)

continued on p. III - 86

TABLE 1 (continued)

Article	Grades	Location	Sample size	Duration	Design	Effect Sizes		
						By achievement	By subject	Total
Breidenstine, 1936	7–9	Soudersburg, PA	11 schools, 860 students	1 yr.	Compared students in 4 AG, 7 hetero schools matched on IQ.		Composite achievement	−.19
Purdom, 1929	9	—[a]	700 students	1 sem.	Matched students in AG, hetero English and algebra classes compared in achievement.	HI −.02 AV −.08 LO +.07	English −.02 Algebra .00	+.01
Postlethwaite & Denton, 1978; Newbold, 1977	5–7	Britain	1 school, 450 students	2 yrs.	Students within one secondary school assigned to streamed or unstreamed halls. Achievement assessed on national examinations.	HI (0) AV (0) LO (0)	Math (0) English (0) Social studies (0) French (0)	(0)
Bachman, 1968	7	Portland, OR	15 schools, 23 teachers, 404 students	1 yr.	Math classes in schools using AG compared to hetero classes, controlling for IQ.		Math (0)	(0)
Kline, 1964	9–12	St. Louis, MO	4 schools	4 yrs.	Retrospective study of successive cohorts of students, one in 3- or 4-group AG, one hetero, in 4 schools. Compared on standardized tests after 4 years of AG or hetero placement.	V.HI −.02 HI +.08 AV .00 LO −.02	Reading −.05 Language +.07 Math + .01	+.01

Stoakes, 1964	7	Cedar Rapids, IA	3 schools	1 yr.	Matched mentally advanced and slow-learning students compared in schools using AG or hetero assignment. Compared on standardized tests.	HI (0) LO (0)	Reading (0) English (0) Math (0)	(0)
Martin, 1959	6–8	Nashville, TN	3 schools	2 yrs.	Retrospective study compared gains on Stanford Achievement Tests for 2 AG and 1 hetero school from Grades 6–8.	HI (0) AV (0) LO (0)	Reading (0) Language (0) Math (0)	(0)
Chiotti, 1961	9	Issaquah, WA	3 schools	1 yr.	Matched students in 3-group AG and hetero schools compared in math achievement.	HI +.14 AV +.06 LO +.35	Math +.18	+.18
Fowlkes, 1931	7	Glendale, CA	2 schools	1 sem.	Students in school using 3-group AG based on IQ matched with students in hetero school. Compared gains on Stanford Achievement Tests.	HI −.45 AV −.18 LO −.05	Reading −.04 Language −.17 Math −.17 Social studies −.21	−.20
Cochran, 1961	8	Kalamazoo, MI	1 school	1 yr.	Compared students grouped separately for English, math, to previous year (hetero) students matched in IQ, age, sex, achievement.	HI (0) AV (0) LO (0)	Math (0) English (0)	(0)

Note. AG = ability grouping; hetero = heterogeneous assignment; HI = high achieving students; AV = average achieving students; LO = low achieving students.

[a] Not available.

grouping per se. For example, a secondary school that used heterogeneous grouping might have a staff, principal, or community more concerned about equity, affective development, or other goals than would a "matched" school that used ability grouping. However, several of the correlation studies used very large samples and longitudinal designs, and these provide important additional information not obtainable from the typically smaller and shorter experimental studies.

Within each category, studies are listed in descending order of sample size. All other things being equal, therefore, studies near the top of Table 1 should be considered as better evidence of the effects of ability grouping that studies near the end of the table. However, the nature and quality of the studies are discussed in more detail in the following sections.

Overall Findings

Across the 29 studies listed in Table 1, the effects of ability grouping on student achievement are essentially zero. The median effect size (ES) for the 20 studies from which effect sizes could be estimated was −.02, and none of the 9 additional studies found statistically significant effects. Counting the studies with nonsignificant differences as though they had effect sizes of .00, the median effect size for all 29 studies would be .00. Results from the 15 randomized and matched experimental studies were not much different; the median effect size was −.06 for the 13 studies from which effect sizes could be estimated. In 9 of these 13 studies (including all 5 of the randomized studies) results favored the heterogeneous groups, but these effects are mostly very small.

There are few consistent patterns in the study findings. Most of the studies involved Grades 7–9, with ninth graders sometimes in junior high schools and sometimes in senior high schools. No apparent trend is discernible within this range. Above the ninth grade the evidence is too sparse for firm conclusions. Lovell (1960) found that high achieving tenth graders performed significantly better in ability-grouped English classes, but there were no effects in biology or algebra and no effects for average or low achievers. In a 4-year study of students in Grades 9–12, Borg (1965) found significant positive effects of ability grouping for average and low achievers in math but no differences in science or for high achievers. Cohorts followed from Grades 7–10 and 8–11 showed no significant differences on any measure for any ability level. In contrast, Thompson (1974), in a study of 11th grade social studies, found the largest effects favoring heterogeneous grouping (ES = −.48), whereas Kline (1964), in another 4-year study of students in Grades 9–12, found no differences.

Twelve of the 29 studies tracked students for all subjects according to one composite ability or achievement measure. The remaining 17 studies grouped students on the basis of performance in one or more specific subjects. However, there were no differences in the outcomes of these different forms of ability grouping. In addition, there were no consistent patterns in terms of the number of ability groups to which students were assigned (the great majority of studies used 3). Study duration had no apparent impact on outcome. Studies that used adjusted gain scores produced the same effects as other studies, and the use of the adjustment of gain scores described above made no difference in outcomes.

There was no discernible pattern of findings with respect to different subjects, with one possible exception. Studies by Marascuilo and McSweeney (1972), Thompson (1974), and Fowlkes (1931) found relatively strong effects favoring heterogeneous grouping in social studies, and three additional studies by Peterson (1966), Martin (1927), and Postlethwaite and Denton (1978) found no differences or slight effects in the same direction. This is not enough evidence to conclusively point to a positive effect of heterogeneous grouping in social studies, but it is important to note that all three of the randomized or matched experimental studies found differences in this direction.

There were no consistent effects according to study location. All four of the British studies found no differences between streamed and unstreamed classes; a large, longitudinal Swedish study by Svensson (1962), not shown in Table 1 because it lacked adequate evidence of initial equality, also found no differences between streamed and unstreamed classes. Urban, suburban, and rural schools had similar outcomes. The one study that involved large numbers of minority students, a randomized experiment in a New York City high school by Ford (1974), found no differences between ability-grouped and heterogeneous math classes.

Studies conducted before 1950 were no more likely than more recent studies to find achievement differences. On this topic, it is interesting to note that experimental-control studies of ability grouping have not been done in recent years. The only study of the 1980s, by Kerckhoff (1986), was done by a sociologist who focused his attention on differences between students in different streams. This study is described in more detail below. Otherwise, the most recent experimental-control comparisons were done in the early 1970s.

Differential Effects According to Achievement Levels

One of the most important questions about ability grouping in secondary schools concerns the degree to which it differentially affects students at different achievement levels. As noted earlier, many researchers and reviewers, particularly those working the sociological tradition, have emphasized the *relative* impact of grouping for different groups of students far more than the average effect for all students.

Twenty-one of the 29 studies presented in Table 1 presented data on the effects of ability grouping on students of different ability levels. Most studies divided their samples into three categories (high, average, and low achievers), but some used two or four categories.

Across the 15 studies from which effect sizes could be computed, the median effect size was +.01 for high achievers, −.08 for average achievers, and −.02 for low achievers. Effects of this size are indistinguishable from zero, and if all the nonsignificant differences found in studies from which effect sizes could not be computed are counted as effect sizes of .00, the median effect size for each level of student becomes .00. In addition, only one of seven studies from which effect sizes could not be computed (Lovell, 1960) found significantly positive effects of ability grouping for high achievers, and none of these studies found significant effects in either direction for average and low achievers. The randomized and matched experimental studies provided slightly more support for the idea that ability grouping has a differential effect; the median effect sizes for high, average, and low achievers were +.05, −.10, and −.06, respectively. It is interesting to note that the study by Borg (1965), which is often cited to support the differential effect of ability grouping on students of different ability levels, in fact provides very weak support for this phenomenon. Across two measures given to members of four 4-year cohorts that principally included secondary years, significant effects favoring ability grouping were found for high achievers in one of eight comparisons, for average achievers in three of eight, and for low achievers in one of eight. Only in a cohort that included Grades 4 to 7 were there significant effects favoring heterogeneous grouping for low achievers.

It might be expected that differential effects of track placement would build over time and that longitudinal studies would show more of a differential impact than 1-year studies. The one multiyear randomized study, by Marascuilo and McSweeney (1972), did find that over a 2-year period, students in the top social studies classes gained slightly more than similar students in heterogeneous classes (ES = +.14), whereas middle (ES = −.37), and low (ES = −.43) groups gained significantly less than their ungrouped counterparts. However, across seven multiyear correlational studies of up to 5 years' duration, not one found a clear pattern of differential effects.

A few studies provided additional information on differential effects of ability grouping by investigating effects of grouping on high or low achievers only. For example, Torgelson (1963) randomly assigned low achieving students in Grades 7–9 to homogeneous or heterogeneous classes. Across several performance measures, the median effect size was +.13 (nonsignificantly favoring ability grouping). Similarly, Borg and Prpich (1966) randomly assigned low achieving 10th graders to ability-grouped or heterogeneous English classes and found that there were no differences in one cohort. In a second cohort, differences favoring ability grouping on a writing measure were found, but there were no differences on eight other measures.

Studies of ability grouping of high achievers are difficult to distinguish from studies of special programs for the gifted. Well-designed studies of programs for the gifted generally find few effects of separate programs for high achievers unless the programs include acceleration (exposure to material usually taught at a higher grade level) (Fox, 1979; Kulik & Kulik, 1984). That is, grouping per se has little effect on the achievement of high achievers. An outstanding illustration of this is a dissertation by Mikkelson (1962), who randomly assigned high achieving seventh and eighth graders to ability-grouped or heterogeneous math classes. The seventh grade homogeneous classes were given enrichment, but the eighth graders were accelerated, skipping to ninth grade algebra. No effects were found for the seventh graders. The accelerated eighth graders, or course, did substantially better than similar students who were not accelerated on an algebra test, and they did no worse on a test of eighth grade math.

Taken together, research comparing ability-grouped to heterogeneous placements provides little support for the proposition that high achievers gain from grouping whereas low achievers lose. However, there is an important limitation to this conclusion. In most of the studies that compared tracked to untracked grouping plans (including all of the randomized and matched experimental studies), tracked students took different levels of the same courses (e.g., high, average, or low sections of Algebra 1). Yet much of the practical impact of tracking, particularly at the senior high school level, is on determining the nature and number of courses taken in a given area. The experimental studies do not compare students in Algebra 1 to those in Math 9, or students who take 4 years of math to those who take 2. The conclusions drawn in this section are limited, therefore, to the effects of between-class grouping *within the same courses*, and should not be read as indicating a lack of differential effects of tracking as it affects course selection and course requirements.

Other Forms of Ability Grouping

The studies discussed above and summarized in Table 1 evaluated the most common forms of ability grouping in secondary schools—full-time, between-class ability grouping for one or more subjects. However, a few studies have evaluated other grouping plans.

The most widely used form of grouping in elementary schools, within-class ability grouping, has also been evaluated in a few studies involving middle and junior high schools. Campbell (1965) compared the use of three math groups within the class to heterogeneous assignment in two Kansas City junior high schools. There were no differences between the two programs in achievement. Harrah (1956) compared five types of within-class grouping in Grades 7–9 in West Virginia and found ability grouping to be no more successful than other grouping methods. Note that these findings conflict with those of studies of within-class ability grouping in mathematics in the upper elementary grades, which tended to support the use of math groups (Slavin, 1987).

Vakos (1969) evaluated the use of a combination of heterogeneous and homogeneous instruction in 11th grade social studies classes in Minneapolis. Students were grouped by ability 2 days each week, but heterogeneously grouped the other 3 days. No achievement differences were found. Zweibelson, Bahnmuller, and Lyman (1965) evaluated a similar mixed approach to teaching ninth grade social studies in New Rochelle, New York, and also found no achievement differences. Chiotti (1961) compared a flexible plan for grouping junior high school students across grade lines for mathematics to both ability-grouped and heterogeneous grouping plans, and again found no differences in achievement. A cross-grade grouping arrangement similar to the Joplin Plan (Slavin, 1987) was compared to within-class grouping in reading by Chismar (1971) in Grades 4–8. Significantly positive effects of this program were found in Grades 4 and 7 but not 5, 6, and 8.

Reconciling Track/No-Track and High-Track/Low-Track Studies

As noted earlier, two very different traditions of research have dominated research on ability grouping. One involves comparisons of ability-grouped to heterogeneous placements. The other involves comparisons of the progress made by students in different ability groups or tracks. Whereas there has been little experimental research comparing ability-grouped to heterogeneous placements since the early 1970s, research comparing the achievement of students in different tracks largely began in the 1970s and continues to the present.

The findings of high-track/low-track studies of ability grouping conflict with those emphasized in this review in that they generally find that even after controlling for IQ, socioeconomic status, pretests, and other measures, students in high tracks gain significantly more in achievement than do students in low tracks, especially in mathematics (see Gamoran & Berends, 1987, for a review). How can these findings be reconciled with those of the experimental studies?

One important difference between experimental and correlational studies of ability grouping is that, as mentioned earlier, correlational studies (especially at the senior high school level) often include not only the effects of being in a high, average, or low class, but also the effects of differential course taking. Students in academic tracks may score better than those in general or vocational classes because they take more courses or more advanced courses. The experimental studies comparing grouped and ungrouped classes are all studies of grouping per se, holding course taking and other factors constant. The correlational studies examine tracking as it is in practice, where track placement implies differences in course requirements, course taking patterns, and so on. Also, experimental track versus no-track studies are rare beyond the ninth grade, whereas most correlational studies comparing students in high versus low tracks involve senior high schools. The lack of track versus no-track studies at the senior high school level is hardly surprising given the nearly universal use of some form of tracking at that level. However, tracking usually has a different meaning in senior than in junior high school. Whereas junior high school tracking mostly involves different levels of courses (e.g., high English vs. low English), senior high tracking is more likely to involve completely different patterns of coursework (e.g., metal shop vs. French III). Also, the problem of dropouts becomes serious in senior high school; a study of 12th graders unavoidably excludes the students who may have suffered most from being in the low track and left school (see Gamoran, 1987). This could reduce observed differences between high- and low-track students.

There is limited evidence, however, that differences in course taking or grade level account for the different conclusions of the track/no-track and high-track/low-track studies. Four-year longitudinal studies in U.S. senior high schools by Kline (1964) and Borg (1965) found no differential effects of track placement for high, average, and low achievers (as compared to similar students in untracked placements). Presumably, course-taking patterns in these senior high school studies varied by track. A correlational study by Alexander and Cook (1982) found that although taking more courses in senior high school did increase achievement (controlling for background factors), different course-taking patterns in different tracks did not account for track differences in achievement. Gamoran (1987) found that track effects on math and science achievement were explained in part by the fact that students in the academic tracks take more math and science courses and, in particular, more *advanced* courses in these areas. However, no such patterns were seen on reading, vocabulary, writing, or civics achievement measures. Gamoran noted the difficulty of disentangling track and course taking, which are highly correlated in math and science (and, of course, both track and course taking are strongly correlated with ability, socioeconomic status, and other factors). It is certainly logical to expect correlational studies of senior high school tracking to find different effects of different track placements because of different course-taking patterns, but because of confounding of tracking, course-taking, and student background factors, that is difficult to determine conclusively.

Another likely explanation for different findings of track/no-track and high-track/low-track studies involves the difficulty of statistically controlling for large differences. Students in higher tracks tend to achieve at much higher levels than those in lower tracks (both before and after taking secondary courses), and statistically controlling for these differences is probably not sufficient to completely remove the influence of ability or prior performance on later achievement. Further, studies in higher tracks are also likely to be higher in such attributes as motivation, internal locus of control, academic self-esteem, and effort, factors that are not likely to be controlled in correlational studies.

To understand the difficulty of controlling for large initial differences between students, imagine an experiment in which a new instructional method was to be evaluated. The experimenter selects a group of students who have high test scores and high IQ scores and are nominated by their teachers as being hard working, motivated, and college material. This group becomes the experimental group, and the remaining students serve as the control group. To control for the differences between the groups, prior composite achievement and socioeconomic status are used as covariates or control variables.

In such an experiment, no one would doubt that regardless of the true effectiveness of the innovative treatment, the experimental group would score far better than the control group, even controlling for prior achievement and socioeconomic status. No journal or dissertation committee would accept such a study. Yet this "experiment" is essentially what is being done when researchers compare students in different tracks. When there are significant pretest differences, use of statistical controls through analysis of covariance or regression are considered inadequate to equate the groups. Most often, the statistical controls will undercontrol for true differences (Lord, 1960; Reichardt, 1979). Yet high- and low-track students usually differ in pretests or IQ by one to two standard deviations, an enormous systematic difference for which no statistical procedure can adequately control.

The only study that compared both tracked to untracked schools and high-track to low-track students was a 5-year longitudinal study by Kerckhoff (1986) in Britain. This study illustrates the problem of controlling for large differences. For example, in mathematics, boys in the high track of three-group ability grouping programs gained about 11 z score points from a test given at age 11 to one given at age 16, whereas students in a remedial track gained 18 z score points. Yet the regression coefficient comparing the high-track to ungrouped students was +2.34, indicating performance about 42% of a standard deviation above "predicted" performance. In contrast, the remedial-track boys had a regression coefficient (in comparison to ungrouped students) of −.72, indicating performance about 13% of a standard deviation below "predicted" performance, despite the fact that the remedial students actually gained more than the top-track students. The reason for this is that the remedial students started out (at age 11) scoring 1.64 standard deviations below the ungrouped students, whereas top-track students started out 1.02 standard deviations above the ungrouped students, a total difference between top-track and remedial students of 2.66. No regression or analysis of covariance can adequately control for such large pretest differences. Because of unreliability in the measures and less-than-perfect within-group correlations of pre- and posttests, "predicted" scores based on pretests and other covariates will (other things being equal) be too low for high achievers and too high for low achievers.

Another factor that can contribute to overestimates of the effects of curriculum track on achievement in studies lacking heterogeneous comparison groups is fan spread. Put simply, high achievers usually gain more per year than do low achievers, so over time the gap between high and low achievers grows. This increasing gap cannot be unambiguously ascribed to ability grouping or other school practices, as it occurs under virtually all circumstances. A student who is performing at the 16th percentile in the 6th grade and is still at the 16th percentile in 12th grade will be further "behind" the 12th grade mean in grade equivalents, for example (Coleman & Karweit, 1972).

An additional factor that can contribute to spurious findings indicating a benefit of being in the high track is that factors other than test scores factor into placement decisions. For example, a study by Balow (1964) found that on math tests not used for group placement, there was enormous overlap between students in supposedly homogeneous seventh-grade math classes. More than 72% of the students scored between the lowest score in the top group and the highest score in the bottom group. Among these students in the "area of overlap," students who were in the top group gained the most in math achievement over the course of the year, whereas those in the low group gained the least.

On its surface this study provides support to the "self-fulfilling prophecy" argument. Yet consider what is going on. Imaging two students with identical scores, one assigned to the high group and one to the low group. Why were they so assigned? Random error is a possibility but all the systematic possibilities weigh in the direction of higher performance for the student assigned to the high group. Because teacher judgement was involved, teachers may have accurate knowledge of student motivation, self-esteem, behavior, or other factors to enable them to predict who will do well and who will not. The actual assignments were done on different tests than those used in the Balow study; it is likely that students who scored low on Balow's pretests but were put in the high groups scored high on the test used for placement, and then regressed to a higher mean on Balow's posttest.

What this discussion is meant to convey is not that different tracks do or do not have a differential impact on student achievement, but that comparisons of students in existing tracks cannot tell us one way or another. To learn about the differential impacts of track placement, there are two types of research that might be done. One would be to randomly assign students at the margin to different tracks, something that has never been done. The other is to compare similar students randomly assigned to ability-grouped or ungrouped systems. This has been done several times, and, as noted earlier in this review, there is no clear trend indicating that students in high-track classes learn any more than high achieving students in heterogeneous classes, or that students in low-track classes learn any less than low achieving students in heterogeneous classes.

Why Is Ability Grouping Ineffective?

The evidence summarized in Table 1 and discussed in this review is generally consistent with the conclusions of earlier reviews comparing homogeneous and heterogeneous grouping (e.g., Kulik & Kulik, 1982, 1987; Noland, 1985), but runs counter to two quite different kinds of "common sense." On one hand, it is surprising to find that assignment to the low-ability group is not detrimental to student learning. A substantial literature has indicated the low quality of instruction in low groups (e.g., Evertson, 1982; Gamoran, 1989; Oakes, 1985), and a related body of research has documented the negative impact of ability grouping on the motivations and self-esteem of students assigned to low groups (e.g., Cottle, 1974; Schafer & Olexa, 1971; Trimble & Sinclair, 1987). How can the effect of ability grouping on low achieving students be zero, as this review concludes?

On the other hand, another kind of "common sense" would argue that, at least in certain subjects, ability grouping is imperative in secondary schools. How can an 8th grade math teacher teach a class composed of students who are fully ready for algebra and students who are still not firm in subtraction and multiplication? How does an English teacher teach literature and writing to a class in which reading levels range from 3rd to 12th grade? Yet study after study, including randomized experiments of a quality rarely seen in educational research, finds no positive effect of ability grouping in any subject or at any grade level, even for the high achievers most widely assumed to benefit from grouping.

The present review cannot provide definitive answers to these questions. However, it is worthwhile to speculate on them.

One possibility is that the standardized tests used in virtually all of the studies discussed in this review are too insensitive to pick up effects of grouping. This seems particularly plausible in looking at tests of reading, because reading has not generally been taught as such in secondary schools. However, standardized tests of mathematics do have a great deal of face validity and curricular relevance, and these show no more consistent a pattern of outcomes. Marascuilo and McSweeney (1972) used both teacher-made and standardized measures of social studies achievement and found similar results with each.

Another possibility is that it simply does not matter whom students sit next to in a secondary class. Secondary teachers use a very narrow range of teaching methods, overwhelmingly using some form of lecture or discussion (Goodlad, 1983). In this setting, the direct impact of students on one another may be minimal. If this is so, then any impacts of ability grouping on students would have to be mediated by teacher characteristics or behaviors or by student perceptions and motivations.

Studies contrasting teaching behaviors in high- and low-track classes usually find that the low tracks have a slower pace of instruction and lower time on-task (e.g., Evertson, 1982; Oakes, 1982). Yet, as noted earlier, the meaning and impact of these differences are not self-evident. It may be that a slower pace of instruction is appropriate with lower-achieving students, or that pace is relatively unimportant because a higher pace with lower mastery is essentially equivalent to a lower pace with higher mastery. Higher time on-task should certainly be related to higher achievement (Brophy & Good, 1986), but the comparisons of time on task between high and low tracks are misleading. What would be important to compare is time on task *for low achievers* in homogeneous and heterogeneous classes, because low achievers may simply be off-task more than high achievers regardless of their class placement. In this regard, it is important to note that Evertson, Sanford, and Emmer (1981) found time on-task to be lower in extremely heterogeneous junior high school classes than in less heterogeneous ones because teachers had difficulty managing the more heterogeneous classes.

The lesson to be drawn from research on ability grouping may be that unless teaching methods are systematically changed, school organization has little impact on student achievement. This conclusion would be consistent with the equally puzzling finding that substantial reductions in class size have little impact on achievement (Slavin, 1989); if teachers continue to use some form of lecture/discussion/seatwork/quiz, then it may matter very little in the aggregate which or how many students the teachers are facing. In contrast, forms of ability grouping that were found to make a difference in the upper elementary grades—the Joplin Plan (cross-grade grouping in reading to allow for whole-class instruction) and within-class grouping in mathematics (Slavin, 1987)—both significantly change time allocations and instructional activities within the classroom.

Alternatives to Ability Grouping

If the effects of ability grouping on student achievement are zero, then there is little reason to maintain the practice. As noted earlier in this article, arguments in favor of ability grouping depend on assumptions about the effectiveness of grouping, at least for high achievers. In the absence of any evidence of effectiveness, these arguments cannot be sustained.

Yet there is also no evidence that simply moving away from traditional ability grouping practices will in itself enhance student achievement, and there are legitimate concerns expressed by teachers and others about the practical difficulties of teaching extremely heterogeneous classes as the secondary level. How can schools moving away from traditional ability grouping use this opportunity to contribute to student achievement?

One alternative to ability grouping pften proposed (e.g., Oakes, 1985) is the use of cooperative learning methods, which involve students working in small, heterogeneous learning groups. Research on cooperative learning consistently finds positive effects of these methods if they incorporate two major elements: group goals and individual accountability (Slavin, 1990). That is, the cooperating groups must be rewarded or recognized on the basis of the sum or average of individual learning performances. Cooperative learning methods of this kind have been used successfully at all grade levels, but there is less research on them in Grades 10–12 than in Grades 2–9 (see Newmann & Thompson, 1987). Cooperative learning methods have also had consistently positive impacts on such outcomes as self-esteem, race relations, acceptance of mainstreamed academically handicapped students, and ability to work cooperatively (Slavin, 1990).

One category of cooperative learning methods may be particularly useful in middle schools moving toward heterogeneous class assignments. These methods are Cooperative Integrated Reading and Composition (Stevens, Madden, Slavin, & Farnish, 1987) and Team Assisted Individualization—Mathematics (Slavin & Karweit, 1985; Slavin, Madden, & Leavey, 1984). Both of these methods are designed to accommodate a wide range of student performance levels in one classroom, using both homogeneous and heterogeneous within-class groupings. These programs have been successfully researched in Grades 3–6 but are often used up to the eighth grade level.

Other alternatives to between-class ability grouping have also been found to be successful in the upper elementary grades (see Slavin, 1987) and could probably be effective in middle schools as well. These include within-class ability grouping in mathematics (e.g., teaching two or three math groups within a heterogeneous class) and the Joplin Plan in reading. The Joplin Plan involves regrouping students for reading across grade levels but according to reading level, so that no within-class reading groups are necessary. However, although these alternatives to between-class grouping are promising because of their success in the upper elementary grades, the few studies of within-class ability grouping at the junior high school level have not found this practice to be effective (Campbell, 1965; Harrah, 1956), and the one middle school study of the Joplin Plan found only inconsistent positive effects (Chismar, 1971). (For descriptions of secondary schools implementing alternatives to traditional ability grouping, see Slavin, Braddock, Hall, & Petza, 1989.)

Limitations of This Review

It is important to note several limitations of the present review. Perhaps the most important is that in none of the studies reviewed here were there systematic observations made of teaching and learning. Observational studies and outcome studies have proceeded on parallel tracks; it would be important to be able to relate evidence of outcomes to changes in teacher behaviors or classroom characteristics. In particular, it would be important to know the degree to which teachers in ability-grouped schools actually differentiate instruction. For example, are teacher of high-track classes more likely to provide enrichment (e.g., greater depth on the same objectives) or acceleration (e.g., coverage of more material usually taught at a later grade level)? How do teachers of low-track classes adapt instruction to the needs of their students? How do teachers of untracked, heterogeneous classes accommodate the wide range of performance levels in their classes? What level and pace of instruction is provided in untracked, heterogeneous classes? Most important, how do variations from teacher to teacher in instructional behaviors in high, low, and heterogeneous classes relate to the outcomes of ability grouping for students of different ability levels?

Another limitation, mentioned earlier, is that almost all studies reviewed here used standardized tests of unknown relationship to what was actually taught. It may be, for example, that positive effects of ability grouping for high achievers could be missed by standardized tests because what these students are getting is enrichment or higher-order skills not assessed on the standardized measures, or that negative effects for low achievers are missed because teachers of low-track classes are hammering away at the minimum skills that are assessed on the standardized tests but ignoring other content. Future research on ability grouping needs to closely examine possible outcomes of grouping on more broadly based and sensitive measures.

A third limitation is the age of most of the studies reviewed. It is possible that schools, students, or ability grouping have changed enough since the 1960s and 1970s to make conclusions from these and older studies tenuous.

As noted earlier, the results reported in this review mainly concern the effects of grouping per se, with little regard for the effects of tracking on such factors as course taking. Effects of tracking on differential course taking are most important in senior high schools. There is a need for additional research comparing tracked to untracked situations at the senior high school level, particularly research designed to disentangle the effects of tracking from those of differential course taking.

In addition, it would add greatly to the understanding of ability grouping in secondary schools to have evaluations or even descriptions of a wider range of alternatives to traditional ability grouping. The few studies of within-class grouping, cross-grade groupings, and flexible grouping plans are not nearly adequate to explore alternatives. Cooperative learning, often proposed as an alternative to ability grouping, has frequently been found to increase student achievement in ability-grouped as well as ungrouped secondary classes (Newmann & Thompson, 1987; Slavin, 1990), yet no study has compared cooperative learning in heterogeneous classes to traditional instruction in homogeneous ones. Descriptions of creative alternatives to ability grouping currently exist only at the anecdotal level (Slavin et al., 1989).

Conclusions

Although there are limitations to the scope of this review and to the studies on which it is based, there are several conclusions that can be advanced with some confidence. These are as follows:

1. Comprehensive between-class ability grouping plans have little or no effect on the achievement of secondary students, at least as measured by standardized tests. This conclusion is most strongly supported in Grades 7–9, but the more limited evidence that does exist from studies in Grades 10–12 also fails to support any effect of ability grouping.
2. Different forms of ability grouping are equally ineffective.
3. Ability grouping is equally ineffective in all subjects, except that there may be a negative effect of ability grouping in social studies.
4. Assigning students to different levels of the same course has no consistent positive or negative effects on students of high, average, or low ability.

For the narrow but extremely important purpose of determining the impact of ability grouping on standardized achievement measures, the studies reviewed here are exemplary. Six randomly assigned individual students to ability-grouped or heterogeneous classes, and nine more individually matched students and then assigned them to one or the other grouping plan. Many of the studies followed students for 2 or more years. If there had been any true effect of ability grouping on student achievement, this set of studies would surely have detected it.

For practitioners, the findings summarized above mean that decisions about whether or not to ability group must be made on bases other than likely effects on achievement. Given the antidemocratic, antiegalitarian nature of ability grouping, the burden of proof should be on those who would group rather than those who favor heterogeneous grouping, and in the absence of evidence that grouping is beneficial, it is hard to justify continuation of the practice. The possibility that students in the low groups are at risk for delinquency, dropout, and other social problems (e.g., Rosenbaum, 1980) should also weigh against the use of ability grouping. Yet schools and districts moving toward heterogeneous grouping have little basis for expecting that abolishing ability grouping will in itself significantly accelerate student achievement unless they also undertake changes in curriculum or instruction likely to improve actual teaching.

There is much research still to be done to understand the effects of ability grouping in secondary schools on students achievement. Studies using more sensitive achievement measures, studies of grouping at Grades 10–12, studies of a broader range of alternatives to grouping, and studies relating observations to outcomes of grouping are areas of particular need. Enough research has been done comparing the effects of ability grouping on standardized achievement tests for students assigned to high, middle, and low tracks, at least up through the ninth grade. It is time to move beyond these simple comparisons to consider more fully how secondary schools can adapt instruction to the needs of a heterogeneous student body.

References

Adamson, D. P. (1971). *Differented multi-track grouping vs. uni-track educational grouping in mathematics.* Unpublished doctoral dissertation, Brigham Young University.

Alexander, K. L., & Cook, M. A. (1982). Curricula and coursework: A surprise ending to a familiar story. *American Sociological Review, 47,* 626–640.

Alexander, K. L., Cook, M. A., & McDill, E. L. (1978). Curriculum tracking and educational stratification. *American Sociological Review, 43,* 47–66.

Atkinson, J. W., & O'Connor, P. (1963). *Effects of ability grouping in schools related to individual differences in achievement-related motivation* (Final Report, Cooperative Research Project, No. OE-2-10-024). Washington, DC: U.S. Dept. of Health, Education, and Welfare.

Bachman, A. M. (1968). *Factors related to the achievement of junior high school students in mathematics.* Unpublished doctoral dissertation, University of Oregon.

Bailey, H. P. (1968). A study of the effectiveness of ability grouping on success in first-year algebra. *Dissertation Abstracts International, 28,* 3061A. (University Microfilms No. 68-1249)

Balow, I. (1964). The effects of "homogeneous" grouping in seventh grade arithmetic. *Arithmetic Teacher, 12,* 186–191.

Bangert-Drowns, R. L. (1986). Review of developments in meta-analytic method. *Psychological Bulletin, 99,* 388–399.

Barton, D. P. (1964). An evaluation of ability grouping in ninth grade English. *Dissertation Abstracts International, 25,* 1731. (University Microfilms No. 64-9939)

Bicak, L. J. (1962). *Achievement in eighth grade science by heterogeneous and homogeneous classes.* Unpublished doctoral dissertation, University of Minnesota.

Billett, R. O. (1928). A controlled experiment to determine the advantages of homogeneous grouping. *Educational Research Bulletin, 7,* 139.

Borg, W., (1965). Ability grouping in the public schools. *The Journal of Experimental Education, 34,* 1–97.

Borg, W., & Prpich, T. (1966). Grouping of slow learning high school pupils. *Journal of Secondary Education, 41,* 231–238.

Braddock, J. H. (1990). *Tracking: Implications for student race-ethnic subgroups* (Technical Report No. 1). Baltimore: Center for Research on Effective Schooling for Disadvantaged Students.

Breidenstine, A. G. (1936). The educational achievement of pupils in differentiated and undifferentiated groups. *Journal of Experimental Education, 5,* 91–135.

Brophy, J. E., & Good, T. C. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 328–375). New York: Macmillan.

Campbell, A. L. (1965). A comparison of the effectiveness of two methods of class organization for the teaching of arithmetic in junior high school. *Dissertation Abstracts International, 26,* 813–814. (University Microfilms No. 65-6726)

Chiotti, J. F. (1961). *A progress comparison of ninth grade students in mathematics from three school districts in the state of Washington with varied methods of grouping.* Unpublished doctoral dissertation, University of Northern Colorado.

Chismar, M. H. (1971). *A study of the effectiveness of cross-level grouping of middle school* *under-achievers for reading instruction.* Unpublished doctoral dissertation, Kent State University.

Cochran, J. (1961). Grouping students in junior high school. *Educational Leadership, 18,* 414–419.

Coleman, J. S., & Karweit, N. L. (1972). *Information systems and performance measures in schools.* Englewood Cliffs, NJ: Educational Technology Publications.

Cottle, T. J. (1974). What tracking did to Ollie Taylor. *Social Policy, 5,* 21–24.

Crespo, M., & Michelna, J. (1981). Streaming, absenteeism, and dropping out. *Canadian Journal of Education, 6,* 40–55.

Dar, Y., & Resh, N. (1986). Classroom intellectual composition and academic achievement. *American Educational Research Journal, 23,* 357–374.

Drews, E. M. (1963). *Student abilities, grouping patterns, and classroom interaction* (Cooperative Research Project No. 608). Washington, DC: U.S. Department of Health, Education, and Welfare.

Esposito, D. (1973). Homogeneous and heterogeneous ability grouping: Principal findings and implications for evaluating and designing more effective educational environments. *Review of Educational Research, 43,* 163–179.

Evertson, C. M. (1982). Differences in instructional activities in higher- and lower-achieving junior high English and math classes. *Elementary School Journal, 82,* 329–350.

Evertson, C., Sanford, J., & Emmer, E. (1981). Effects of class heterogeneity in junior high school. *American Educational Research Journal, 18,* 219–232.

Feldhusen, J. F. (1989). Synthesis of research on gifted youth. *Educational Leadership, 46*(6), 6–11.

Ferri, E. (1971). *Streaming: Two years later.* London: National Foundation for Educational Research in England and Wales.

Fick, W. W. (1963). The effectiveness of ability grouping in seventh grade core classes. *Dissertation Abstracts International, 23,* 2753.

Findley, W. G., & Bryan, M. (1971) *Ability grouping: 1970 status, impact, and alternatives.* Athens: Center for Educational Improvement, University of Georgia. (ERIC Document Reproduction Service No. ED 060 595)

Fogelman, K., Essen, J., & Tibbenham, A. (1978). Ability-grouping in secondary schools and attainment. *Educational Studies, 4,* 201–212.

Ford, S. (1974). *Grouping in mathematics: Effects on achievement and learning environment.* Unpublished doctoral dissertation, Yeshiva Univesity.

Fowlkes, J. G. (1931). Homogeneous or heterogeneous group—which? *The Nation's Schools, 8,* 74–78.

Fox, L. H. (1979). Programs for the gifted and talented: An overview. In A. H. Passow (Ed.), *The gifted and talented: Their education and development* (pp. 104–126). Chicago: University of Chicago Press.

Gamoran, A. (1987). The stratification of high school learning opportunities. *Sociology of Education, 60,* 135–155.

Gamoran, A. (1989). Measuring curriculum differentiation. *American Journal of Education, 97,* 129–143.

Gamoran, A., & Berends, M. (1987). The effects of stratification in secondary schools: Synthesis of survey and ethnographic research. *Review of Educational Research, 57,* 415–435.

Gamoran, A., & Mare, R. D. (1989). Secondary school tracking and educational inequality: Compensation, reinforcement, or neutrality? *American Journal of Sociology, 94,* 1146–1183.

Glass, G., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research.* Beverly Hills, CA: Sage.

Good, T., & Marshall, S. (1984). Do students learn more in heterogeneous or homogeneous groups? In P. Peterson, L. C. Wilkinson, & M. Hallinan (Eds.), *The social context of instruction: Group organization and group processes* (pp. 15–38). New York: Academic Press.

Goodlad, J. I. (1983). *A place called school.* New York: McGraw-Hill.

Harrah, D. D. (1956). A study of the effectiveness of five kinds of grouping in the classroom. *Dissertation Abstracts International, 17,* 715. (University Microfilms No. 56-1114)

Heathers, G. (1969). Grouping. In R. Ebel (Ed.), *Encyclopedia of educational research* (4th ed., pp. 559–570). New York: Macmillan.

Holy, T. C., & Sutton, D. H. (1930). Ability grouping in the ninth grade. *Educational Research Bulletin, 9,* 419–422.

Jones, J. D., Erickson, E. L., & Crowell, R. (1972). Increasing the gap between Whites and Blacks: Tracking as a contribution source. *Education and Urban Society, 4,* 339–349.

Kerckhoff, A. C. (1986). Effects of ability grouping in British secondary schools. *American Sociological Review, 51,* 842–858.

Kline, R. E. (1964). A longitudinal study of the effectiveness of the track plan in the secondary schools of a metropolitan community. *Dissertation Abstracts International, 25,* 324. (University Microfilms No. 64-4257)

Kulik, C. L., & Kulik, J. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. *American Educational Research Journal, 19,* 415–428.

Kulik, C., & Kulik, J. (1984). Effects of accelerated instruction on students. *Review of Educational Research, 54,* 409–425.

Kulik, J. A., & Kulik, C. L. (1987). Effects of ability grouping on student achievement. *Equity and Excellence, 23,* 22–30.

Lord, F. M. (1960). Large-sample covariance analysis when the control variable is fallible. *Journal of the American Statistical Association, 55,* 307–321.

Lovell, J. T. (1960). The Bay High School experiment. *Educational Leadership, 17,* 383–387.

Marascuilo, L. A., & McSweeney, M. (1972). Tracking and minority student attitudes and performance. *Urban Education, 6,* 303–319.

Martin, W. B. (1959). Effects of ability grouping on junior high school achievement. *Dissertation Abstracts International, 20,* 2810. (University Microfilms No. 59-1108)

Martin, W. H. (1927). *The results of homogeneous grouping in the junior high school.* Unpublished doctoral dissertation, Yale University.

McPartland, J. M., Coldiron, J. R., & Braddock, J. H. (1987). *School structures and classroom practices in elementary, middle, and secondary schools.* Baltimore: Johns Hopkins University, Center for Research on Elementary and Middle Schools.

Mikkelson, J. E. (1962). *An experimental study of selective grouping and acceleration in junior high school mathematics.* Unpublished doctoral dissertation, University of Minnesota.

Millman, J., & Johnson, M. (1964). Relation of section variance to achievement gains in English and mathematics in grades 7 and 8. *American Educational Research Journal, 1,* 47–51.

Newbold, D. (1977). *Ability-grouping, the Banbury enquiry.* Slough, England: National Foundation for Educational Research in England and Wales.

Newmann, F. M., & Thompson, J. (1987). *Effects of cooperative learning on achievement in secondary schools: A summary of research.* Madison: University of Wisconsin, National Center on Effective Secondary Schools.

Noland, T. K. (1985). *The effects of ability grouping: A meta-analysis of research findings.* Unpublished doctoral dissertation, University of Colorado.

Oakes, J. (1982). The reproduction of inequity: The content of secondary school tracking. *Urban Review, 14,* 107–120.

Oakes, J. (1985). *Keeping track: How schools structure inequality.* New Haven, CT: Yale University Press.

Persell, C. H. (1977). *Education and inequality: A theoretical and empirical synthesis.* New York: Free Press.

Peterson, R. L. (1966). *An experimental study of the effects of ability grouping in grades 7 and 8.* Unpublished doctoral dissertation, University of Minnesota.

Platz, E. F. (1965). The effectiveness of ability grouping in general science classes. *Dissertation Abstracts International, 26,* 1459A–1460A. (University Microfilms No. 65-6914)

Postlethwaite, K., & Denton, C. (1978). *Streams for the future?* Slough, England: National Foundation for Educational Research.

Purdom, T. L. (1929). *The value of homogeneous grouping.* Baltimore: Warwick & York.

Reichardt, C. S. (1979). The statistical analysis of data from nonequivalent group designs. In T. C. Cook & D. T. Campbell (Eds.), *Quasi-experimentation: Design and analysis issues for field settings* (pp. 147–205). Chicago: Rand McNally.

Rosenbaum, J. E. (1976). *Making inequality: The hidden curriculum of high school tracking.* New York: Wiley.

Rosenbaum, J. E. (1978). The structure of opportunity in school. *Social Forces, 57,* 236–256.

Rosenbaum, J. E. (1980). Social implications of educational grouping. *Review of Research in Education, 8,* 361–401.

Schafer, W., & Olexa, C. (1971). *Tracking and opportunity: The locking-out process and beyond.* Scranton, PA: Chandler.

Slavin, R. E. (1986). Best evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher, 15*(9), 5–11.

Slavin, R. E. (1987). Ability grouping and student achievement in elementary schools: A best-evidence synthesis. *Review of Educational Research, 57,* 347–350.

Slavin, R. E. (1989). Achievement effects of substantial reductions in class size. In R. E. Slavin (Ed.), *School and classroom organization* (pp. 247–257). Hillsdale, NJ: Erlbaum.

Slavin, R. E. (1990). *Cooperative learning: Theory, research, and practice.* Englewood Cliffs, NJ: Prentice-Hall.

Slavin, R. E., Braddock, J. H., Hall, C., & Petza, R. J. (1989). *Alternatives to ability grouping.* Baltimore: Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students.

Slavin, R. E., & Karweit, N. L. (1985). Effects of whole-class, ability grouped, and individualized instruction on mathematics achievement. *American Educational Research Journal, 22,* 351–367.

Slavin, R. E., Madden, N. A., & Leavey, M. B. (1984). Effects of team assisted individualization on mathematics achievement of academically handicapped and non-handicapped students. *Journal of Educational Psychology, 76,* 813–819.

Sorensen, A. B., & Hallinan, M. T. (1986). Effects of ability grouping on growth in academic achievement. *American Educational Research Journal, 23,* 543–557.

Stevens, R. J., Madden, N. A., Slavin, R. E., & Farnish, A. M. (1987). Cooperative integrated reading and composition: Two field experiments. *Reading Research Quarterly, 22,* 433–454.

Stoakes, D. W. (1964). *An educational experiment with the homogeneous grouping of mentally advanced and slow learning students in the junior high school.* Unpublished doctoral dissertation, University of Colorado.

Svensson, N. E. (1962). *Ability grouping and scholastic achievement.* Stockholm: Almqvist & Wirsell.

Thompson, G. W. (1974). The effects of ability grouping upon achievement in eleventh grade American history. *Journal of Experimental Education, 42,* 76–79.

Torgelson, J. W. (1963). *A comparison of homogeneous and heterogeneous grouping for below-average junior high school students.* Unpublished doctoral dissertation, University of Minnesota.

Trimble, K. D., & Sinclair, R. L. (1987). On the wrong track: Ability grouping and the threat to equity. *Equity and Excellence, 23,* 15–21.

Turney, A. H. (1931). The status of ability grouping. *Educational Administration and Supervision, 17,* 21–42, 110–127.

Vakos, H. N. (1969). The effect of part-time grouping on achievement in social studies. *Dissertation Abstracts International, 30,* 2271. (University Microfilms No. 69-20,066)

Vanfossen, B. E., Jones, J. D., & Spade, J. Z. (1987). Curriculum tracking and status maintenance. *Sociology of Education, 60,* 104–122.

Wiatrowski, M., Hansell, S., Massey, C. R., & Wilson, D. L. (1982). Curriculum tracking and delinquency. *American Sociological Review, 47,* 151–160.

Wilcox, J. (1963). A search for the multiple effects of grouping upon the growth and behavior of junior high school pupils. *Dissertation Abstracts International, 24,* 205.

Willcutt, R. E. (1969). Ability grouping by content topics in junior high school mathematics. *Journal of Educational Research, 63,* 152–156.

Zweibelson, I., Bahnmuller, M., & Lyman, L. (1965). Team teaching and flexible grouping in the junior high-school social studies. *Journal of Experimental Education, 34,* 20–32.

Author

ROBERT E. SLAVIN is Co-Director, Elementary School Program, Center for Research on Effective Schooling for Disadvantaged Students, The Johns Hopkins University, 3505 N. Charles St., Baltimore, MD 21218. He specializes in cooperative learning, school and classroom organization, programs for disadvantaged students, and research review.

THE VARIABLE EFFECTS OF HIGH SCHOOL TRACKING*

Adam Gamoran
University of Wisconsin, Madison

The effects of tracking in high schools depend in part on the way tracking is organized: To the extent that the structure of tracking varies across schools, tracking's impact on achievement also varies. I examine four structural characteristics of tracking systems: selectivity, electivity, inclusiveness, and scope. I predict that differences in these characteristics lead to variation in between-track inequality (the achievement gap between tracks) and school productivity (average achievement of students in the school), net of the composition of the student body. In addition, I hypothesize that Catholic schools have less inequality between tracks and higher productivity overall than public schools. I test the hypotheses using data from High School and Beyond, a national survey of high schools and their students. The results show that schools vary significantly in the magnitude of track effects on math achievement, and they differ in net average achievement on both math and verbal tests. Schools with more mobility in their tracking systems produce higher math achievement overall. They also have smaller gaps between tracks in both math and verbal achievement when compared to schools with more rigid tracking systems. Moderately inclusive systems also have less between-track inequality in math; and overall school achievement tends to rise in both subjects as inclusiveness increases. The hypotheses about Catholic schools are also supported, especially for math achievement. The way Catholic schools implement tracking partially accounts for their advantages.

Many writers have suggested that the effects of high school tracking on student achievement vary among schools, but none has offered a compelling theory for why this may occur (Heyns 1974; Hauser, Sewell, and Alwin 1976; Rosenbaum 1984). I use existing knowledge about tracking to develop hypotheses for between-school differences in tracking's effects. Building on the work of Sørensen (1970), I argue that the impact of tracking varies according to the structural characteristics of school tracking systems. I also consider claims that tracking has different effects in public and Catholic schools (Gamoran and Berends 1987; Page and Valli 1990). I test these hypotheses by applying methods of multilevel contextual analysis to data on tracking and achievement in a national sample of high schools.

TRACKING AND STUDENT ACHIEVEMENT

Tracking may affect academic achievement in two ways. First, it may affect the dispersion of achievement, or educational *inequality*. Tracking adds to inequality when placement in a high-status track permits students to gain more than if they had been assigned to a lower track. A key question is whether some forms of tracking induce more inequality between tracks than others.

Second, the particular structure of tracking may influence a school's overall level of achievement.

---

* Direct all correspondence to Adam Gamoran, Department of Sociology, 1180 Observatory Drive, Madison, WI 53706. Earlier versions of this paper were presented in seminars at the University of Wisconsin, the University of Chicago, and the University of Edinburgh, and at the annual meeting of the American Sociological Association in Atlanta, August 1988. I am grateful for comments received at those forums and for additional suggestions from Charles Bidwell, Anthony Bryk, Robert Hauser, Alan Kerckhoff, Michael Olneck, Åage Sørensen, Douglas Willms, and the *ASR* editors and referees. I also appreciate assistance from Dae-dong Hahn in preparing the data set, and Mary Rasmussen and Richard Congdon in implementing the HLM computer program. The research was supported by the National Center on Effective Secondary Schools at the Wisconsin Center for Education Research, which was supported by grant No. G-008690007-90 from the Office of Educational Research and Improvement. A Spencer Fellowship from the National Academy of Education provided additional support. Any opinions, findings, and conclusions expressed in this paper are my own and do not necessarily reflect the views of these agencies or the U.S. Department of Education.

American Sociological Review, 1992, Vol. 57 (December 812–828)

Copyright 1992 by the American Sociological Association. Reprinted by permission of the publisher

or educational *productivity*. Is one type of tracking system more productive than another? If so, then variation in the structure of tracking contributes to between-school variation in achievement.[1] A concern with productivity must be paired with the study of inequality because if certain forms of tracking reduce achievement gaps between tracks, it is essential to know whether this occurs in a context of higher, lower, or the same overall school achievement.

Does Tracking Affect Inequality?

For many years, students, teachers, and field researchers have reported that more learning occurs in higher tracks (Hargreaves 1967; Rosenbaum 1976; Metz 1978; Ball 1981; Burgess 1984; Oakes 1985). Most survey studies corroborate these reports after controlling for students' initial characteristics, including family background, race, gender, and prior achievement (for a review, see Gamoran and Berends 1987). Although the finding has not been universal — Jencks and Brown (1975) and Alexander and Cook (1982) raised doubts — the evidence for differences between tracks seems persuasive in well-specified, carefully controlled analyses using national survey data for Britain, Israel, and the United States (Kerckhoff 1986; Shavit and Featherman 1988; Natriello, Pallas, and Alexander 1989). Even Slavin (1990), who maintained that between-classroom ability grouping in American secondary schools has no effects, acknowledged that broad curriculum tracking probably magnifies inequality in achievement. Gamoran and Mare (1989) showed that this conclusion holds even after taking into account the effects of unmeasured selection variables.

Previous writers have disagreed about whether tracking's differentiating effects vary across schools. Citing case studies, Rosenbaum (1984) argued vehemently that such variation occurs, and he attributed discrepant survey findings in part to differences among tracking systems. In a nationwide study, Oakes (1985) described considerable variation in the characteristics of tracking systems in 25 junior and senior high schools, but she did not examine whether these differences affected the impact of tracking on achievement.

Heyns (1974) reported statistically significant interactions between track positions and dummy variables representing the high schools in her data set. However, because the interactions were relatively small, and because they were not related to school size or location, she estimated an additive model of tracking's effects. Using the same approach, Hauser, Sewell, and Alwin (1976) found no significant between-school differences in track effects in Milwaukee County, Wisconsin. Consequently they, too, estimated an additive model. Both Heyns's data, which were limited to urban schools outside the South, and Hauser, Sewell, and Alwin's Wisconsin data, may be less variable than a national sample. Because of disagreement over the existence and magnitude of between-school differences in track effects, I first test for homogeneity of effects across schools, and then explore the sources of differences that appear.

Are Some Tracking Systems More Productive Than Others?

A more productive tracking system is one that results in higher average achievement than a less productive one. In light of tracking's effects on inequality, this means that, with given proportions of students in the different tracks, a more productive system must have a greater positive effect for high-track students, or a less negative impact for students in low tracks, or some combination.

Clearly, schools differ in their average achievement levels. Even after adjusting for differences in student body composition, some schools appear more productive than others. Although between-school variation typically constitutes less than 20 percent of the total variation in student achievement, that amount is statistically significant (Bryk and Driscoll 1988; Lee and Bryk

---

[1] I focus on variation among types of tracking systems, not on the presence or absence of tracking. Almost all American high schools use some form of tracking (Oakes, Gamoran, and Page 1992), and available survey data do not readily permit comparisons of tracking to no tracking. Previous work simulated changes in inequality and productivity produced by the hypothetical absence of tracking compared to the average tracking system (Gamoran and Mare 1989). In that study, all tracking systems were assumed to operate similarly, and the question of inequality focused on differences between subgroups (black versus white, female versus male, economically advantaged versus disadvantaged), while the question for productivity considered the average gain due to tracking compared to the simulated absence of tracking. In the present study, I am concerned with inequality in achievement between the tracks themselves, and with differences in the school achievement levels associated with different tracking systems.

1989). We do not know whether this variation is affected by the structure of tracking in schools. Gamoran (1987) reported higher overall vocabulary achievement in schools with larger college tracks, but the relation did not hold for math, science, reading, writing, or civics achievement. Moreover, the effect on vocabulary achievement vanished when students' own track positions were taken into account. To date, researchers have not presented a conceptual account for the impact of the structure of tracking on achievement levels in different tracks or in the school as a whole. Such an account must be based on knowledge of how the effects of tracking come about.

How Do Track Effects Occur?

Prior researchers have suggested that tracking influences student achievement through mechanisms of social-psychological and academic differentiation. Observers in Britain and the United States have argued that secondary school stratification polarizes the student body into pro-school and anti-school factions (Hargreaves 1967; Lacey 1970; Metz 1978; Ball 1981; Schwartz 1981). College-bound students conform to the school's demands, while others resist. Polarization may be stimulated by the labeling of students according to track positions (Schwartz, 1981). Teachers and guidance counselors communicate differential expectations to students by encouraging those in college-bound programs more than others (Hargreaves 1967; Heyns 1974; Ball 1981). Peer groups may also encourage polarization — observers and survey researchers have found that students tend to form friendships with others in the same track (Hargreaves 1967; Hauser, Sewell, and Alwin 1976; Ball 1981; Eckert 1989; Hallinan and Williams 1989). Social relations within friendship groups may promote differentiated attitudes and behavior in school. Presumably as a result of these conditions, high-track students often find greater meaning in school work, are more motivated, put forth greater effort, and hold higher expectations for themselves compared to low-track students. All these factors are said to lead to differences in achievement.

No quantitative study has tested these claims by including student behavior, attitudes, and expectations as intervening variables between track position and achievement. In fact, the evidence is inconclusive as to whether high school tracking actually produces such social-psychological differentiation, or whether it simply reflects differences already in place. Many studies have reported differences among tracks in educational expectations, even after controlling for academic plans at the outset of tracking (Rehberg and Rosenthal 1978; Alexander and Cook 1982; Waitrowski, Hansell, Massey, and Wilson 1982; Vanfossen, Jones, and Spade 1987; Berends 1991). The evidence concerning student attitudes and behavior is more ambiguous. Waitrowski et al. (1982) found no track effects on self-esteem, attachment to school, or delinquent behavior. However, Berends's (1991) results supported the polarization hypothesis for academic engagement and discipline problems, and Vanfossen, Jones, and Spade (1987) reported significant track effects on self-esteem and liking for school. Tracking is clearly implicated in the differentiation of students' educational expectations, and possibly students' attitudes and behavior as well. Variation in expectations, attitudes, and behavior may then contribute to variation in achievement.

Besides social-psychological differentiation, tracking also appears to produce differences in students' academic experiences that further differentiate achievement. Students in college-preparatory programs take more academic courses, particularly in math and science (Gamoran 1987; Vanfossen, Jones, and Spade 1987). In many subject areas, they are exposed to more high-status knowledge (Keddie 1971; Burgess 1983, 1984; Oakes 1985; Page 1987, 1991). Teachers in high-track classes present more complex material at a faster pace (Metz 1978; Ball 1981; Oakes 1985), and survey and observational studies have reported a more positive academic climate in high-track classes (Metz 1978; Oakes 1985; Vanfossen, Jones, and Spade 1987). Finally, teachers reputed to be more skillful are disproportionately assigned to high-track classes (Hargreaves 1967; Lacey 1970; Rosenbaum 1976; Ball 1981; Finley 1984). Although these between-track differences are clearly documented, their mediating role in the relation between tracking and achievement is less well established (Gamoran and Berends 1987; Gamoran, Nystrand, Berends, and LePore 1992). Nonetheless, instructional differentiation appears to be another important mechanism underlying track differences in achievement.

STUDENT ACHIEVEMENT AND THE STRUCTURE OF STRATIFICATION

Sørensen (1970) described organizational differentiation in school systems, and his description was elaborated by Rosenbaum (1976, 1984) and Oakes (1985). In Sørensen's scheme, high school tracking is an instance of horizontal differentiation involving curricular differences within grade levels, and vertical differentiation involving status distinctions between academic and nonacademic programs. Not all tracking systems are alike, however. They vary along several structural dimensions, including (1) *selectivity* — the degree of homogeneity within tracks; (2) *electivity* — whether students choose or are assigned to track positions; (3) *inclusiveness* — the subsequent educational opportunities available; and (4) *scope* — the breadth and flexibility of track assignment. How are these structural characteristics related to variation in between-track inequality and school productivity?

Selectivity

Sørensen defined selectivity as the amount of homogeneity created by grouping students according to characteristics relevant for learning. Classes in a highly selective system are more homogeneous than the student body as a whole. Selectivity can also be viewed as the size of the gap between groups — the top group in a highly selective system is much higher on the selection criterion (e.g., ability) than other groups (Gamoran 1984). Thus, selectivity involves both the variance (homogeneity) and the means (levels) of the groups. The extent to which a school's tracks are homogeneous and distinct is a function of two conditions: the initial heterogeneity of the student body, and the policies that distribute students to tracks.

By definition, highly selective tracking systems are elitist — they place high-achieving students together to form homogeneous classes. Tracking tends to be especially visible in highly selective systems, with high academic status awarded to the "cream of the crop." By emphasizing the top track at the expense of other tracks, selectivity probably magnifies between-track variation in students' educational attitudes and expectations. If so, one would expect high selectivity to accentuate between-track differences in achievement.

Moreover, highly selective tracking systems are often characterized by greater between-track variability in students' instructional experiences. Because teachers adjust instruction to student aptitudes, tracks that differ more in initial levels of student performance are likely to vary more in their instructional regimes (Dahloff 1971; Lundgren 1972; Barr and Dreeben 1983), and hence produce wider gaps in achievement.

H_{1a}: The greater the selectivity of a tracking system, the larger the differences between tracks in achievement, when relevant prior characteristics of students are controlled.

At the same time, greater selectivity may lead to higher achievement overall. Many educators maintain that homogeneous classes allow teachers to tailor the curriculum to students' needs (Wilson and Schmits 1978). If there is an instructional advantage to homogeneous grouping, that advantage is likely to be greater when the groups are more homogeneous (Slavin 1987).

H_{1b}: The greater the selectivity of a tracking system, the higher the overall achievement in the school, when the composition of the student body is controlled.

Hypothesis 1b is a prediction about average achievement in the school, and it does not distinguish among the tracks. Taken together, however, Hypotheses 1a and 1b imply that selectivity adds to inequality by raising achievement in high tracks more than in lower tracks.[2]

Electivity

Electivity refers to the extent to which students choose or are assigned to tracks (Sørensen 1970). Several researchers have reported that even when students formally have a choice of tracks, in practice they are highly influenced by school authorities. Students and their parents are urged by teachers, principals, and guidance counselors to make the "right" choices according to their capacities (Cicourel and Kitsuse 1963; Ball 1981; Gamoran 1992).

Nonetheless, many American high school students believe they chose their track positions (Jencks et al. 1972; Jones, Vanfossen, and Spade 1986).[3] These perceptions may be a more important factor underlying track effects on achievement than are the objective circumstances of assignment. Students who believe they selected their programs are more likely to be motivated to perform, regardless of which track they are in. Thus, one may expect less social-psychological differentiation and, consequently, smaller between-track differences in achievement. Because the lower degree of differentiation occurs through more positive attitudes in all tracks, I also predict higher overall school achievement in a more elective system.

---

[2] Sørensen (1970) also noted that schools differ in the criteria used to assign students to programs. A key issue is the extent to which placement relies on cognitive characteristics, e.g., ability or achievement. This issue can be subsumed under selectivity, because when the selection process relies on achievement, more selective systems by definition involve tighter links between cognitive characteristics and track positions.

[3] In a random subsample of the nationally representative High School and Beyond survey, about two-thirds of high-school sophomores said they chose their curricular program (Jones, Vanfossen, and Spade 1986).

H_{2a}: The greater the degree of electivity in a tracking system, the smaller the differences between tracks in achievement, when relevant prior characteristics of students are controlled.

H_{2b}: The greater the degree of electivity in a tracking system, the higher the average achievement in the school, when the composition of the student body is taken into account.

Sørensen (1970), by contrast, suggested that electivity magnifies tracking's effects on achievement. He reasoned that electivity leads to within-track homogeneity of educational aspirations, which in turn strengthens differential peer-group effects and thus increases the differences in achievement.

Inclusiveness

Inclusive tracking systems leave open students' options for future schooling (Sørensen 1970; see also Rosenbaum 1976 and Kilgore 1991). A high school tracking system is more inclusive if it assigns relatively more students to the college-preparatory curriculum. The larger the size of the college-bound track, the more salient it is likely to be — for those who are left out. The stigma of being excluded is greater when a larger proportion of students are included (Page 1991). For example, membership in a noncollege program may incur greater stigma when it consists of the bottom 10 percent of the school's academic hierarchy than when it is the bottom 40 percent. Although an inclusive system is less elitist, it is highly visible and thus stigmatizes those left out of the preferred group.

However, a system characterized by very low inclusiveness also probably raises the salience of the college track. Like high selectivity, low inclusiveness reflects an elitist system, which may increase the degree of social-psychological and instructional differentiation among tracks. Hence, I expect larger differences between tracks in achievement when inclusiveness is very low as well as when it is high. Smaller achievement differences may occur when students are more evenly distributed across tracks.

H_{3a}: Controlling for relevant prior characteristics of students, track differences in achievement are larger when the system is highly inclusive or minimally inclusive, and smaller when inclusiveness is moderate.

The impact of inclusiveness on overall school achievement may also be nonlinear. In general, schools with larger college tracks may have higher average achievement, net of composition and students' track positions, because a large college track reflects greater academic emphasis in the school, which tends to raise achievement for all students regardless of track (Powell, Farrar, and Cohen 1985; Lee and Bryk 1988). This effect, however, probably declines as inclusiveness becomes very high because, as the academic track expands, students who are left out become increasingly stigmatized (Hypothesis 3a), depressing mean achievement. Thus, as the size of the academic track increases, the benefits of inclusiveness may decline.

H_{3b}: Controlling for compositional differences, greater inclusiveness in a tracking system contributes to higher average achievement, but at a declining rate.

This nonlinearity may account for the weak linear effects of size of academic track on student achievement observed in earlier work (Gamoran 1987).

Scope

Sørensen (1970) viewed scope as the extent to which students are located in the same track across subjects. Rosenbaum (1976) added track mobility (movement of students across tracks) to the concept.[4] Oakes (1985) further characterized scope to include "extent" (the proportion of all classes that are tracked), "pervasiveness" (the number of subject areas that are tracked), and

---

[4] Sørensen (1970) distinguished scope from "the rigidity of differentiation . . . the extent to which students may transfer to another group than the one originally assigned to" (1970, p. 363, note 2). Although Sørensen believed this would involve few students, recent data suggest transfers are common, at least as indicated by self-report data (Gamoran 1987). Thus, the permanence of assignments, or track immobility, is considered part of tracking's scope.

"flexibility" (whether track assignments are made for each subject or across all subjects).

Tracking systems with wide scope are likely to be more salient to students than systems with narrower scope. Status distinctions may be more meaningful if they apply to a large share of a student's school day and if they are consistent across subjects. In addition, tracking systems that cover more subjects and allow less mobility are more likely to produce differential friendship networks (Sørensen 1970). The socialization effects of tracking are thus compounded in a system of wide scope, and therefore differences among tracks in achievement may be larger.

Wider scope also means greater between-track variation in students' academic experiences. Students grouped for more subjects and for a longer time period are exposed to more differentiated instruction, thus increasing the net effects of track position on achievement.

H_{4a}: The wider the scope of a tracking system, the larger the differences between tracks in achievement, when relevant prior characteristics of students are controlled.

Also, a tracking system that is inflexible over time and across subjects may result in lower overall achievement in the school, compared to a more flexible system. Failure to adjust assignments for developmental, motivational, or other changes in students' capacities for learning, and failure to recognize differences in students' aptitudes for different subjects, impede the match of instruction to student needs (Slavin 1987). Hence, the differentiating effect of wide scope is likely to occur in a context of lower overall achievement.

H_{4b}: The wider the scope of the tracking system, the lower the average achievement in the school, when student body composition is controlled.

Public Versus Catholic Schools

Prior research has suggested that tracking in Catholic schools differs from tracking in public schools. First, Catholic schools place greater academic demands on students in noncollege tracks, requiring more academic courses and more rigorous classwork, compared to noncollege tracks in public schools (Hoffer, Greeley, and Coleman 1985; Lee and Bryk 1988; Camarena 1990). Hence, the degree of instructional differentiation between tracks may be lower in Catholic schools. Second, an observational study of three Catholic high schools reported that students and teachers hold positive views about assignment to low tracks and are optimistic about the possibility of remediation (Valli 1990). This finding contrasts with the negative attitudes typically found in public schools (e.g., Oakes 1985), and suggests that tracking's impact on social-psychological differentiation may be less in Catholic schools. For these reasons, net achievement gaps between tracks are likely to be smaller in Catholic schools than in public schools.

H_{5a}: Differences between tracks in achievement are smaller in Catholic schools than in public schools, when relevant prior characteristics of students are controlled.

Several studies reported higher average achievement in Catholic schools compared to public schools (Hoffer et al. 1985; Lee and Bryk 1988, 1989; for critiques, see Alexander and Pallas 1985; Willms 1985; Jencks 1985). Part of the Catholic-school advantage may be tied to the way tracking is used — the relatively large size of the academic track, the emphasis on academic work in all tracks, and the less stigmatization of low-track students — all may contribute to higher achievement in Catholic schools (Hoffer et al. 1985; Lee and Bryk 1988; Valli 1990). Hence, I predict higher overall achievement in Catholic schools than in public schools.

H_{5b}: Catholic schools have higher overall achievement, net of compositional differences, compared to public schools. Differences in the structure of tracking account for part of the Catholic-school advantage.

METHODS

These hypotheses describe effects at two levels of analysis: (1) student-level effects on achievement within schools, and (2) school-level effects on between-school differences in the impact of tracking, and on variation in school mean achievement, net of compositional differences. To address both levels of analysis, I use a method called hierarchical linear modeling (HLM) (Bryk and Raudenbush 1992), also known as multilevel contextual analysis (Mason, Wong, and Entwistle 1983; DiPrete and Grusky 1991). HLM estimates equations corresponding to the two levels of analysis. At the student level, achievement within each school is predicted:

$$\begin{aligned}(Achievement)_{ij} &= \beta_{0j} + \beta_{1j}(Track)_{ij} \\ &\quad + \beta_{2j}\,(Background)_{ij} + \varepsilon_{ij}. \quad (1)\end{aligned}$$

In this study, both the intercept, β_{0j}, and the track effect, β_{1j}, are allowed to vary from school to school, while the effects of background variables, β_{2j}, are constrained to be equal across schools.[5] The β coefficients that vary across schools (β_{0j} and β_{1j}) serve as dependent variables in the school-level equations:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(Sector)_j + \gamma_{02}(Structure)_j + \upsilon_{0j}; \quad (2)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(Sector)_j + \gamma_{12}(Structure)_j + \upsilon_{1j}. \quad (3)$$

In equations 2 and 3, "sector" refers to Catholic or public schools, and "structure" stands for the structural characteristics of tracking systems.[6] When the within-school predictors are centered around their grand means, β_{0j} represents school mean achievement adjusted for compositional differences, i.e., net school productivity.[7] Inequality between tracks is reflected in β_{1j}, which is the net achievement gap between tracks in each school. In HLM, equations 1, 2, and 3 are estimated simultaneously, producing maximum-likelihood estimates of the variance components, which are then used to generate the β and γ coefficients (for a more detailed account, see Bryk and Raudenbush 1992).

The HLM approach is superior to traditional techniques for measuring school effects and track effects. For example, one common strategy estimates the entire model at the student level, assigning values of school-level variables to students within schools. This approach uses ordinary least squares (OLS) regression to obtain the track effects, and adds interaction terms to assess the impact of sectoral and structural variation on the effects of tracking:

$$\begin{aligned}(Achievement)_i = {} & \beta_0 + \beta_1(Track)_i \\ & + \beta_2(Background)_i + \beta_3(Sector)_i \\ & + \beta_4(Structure)_i + \beta_5(Track \times Sector)_i \\ & + \beta_6(Track \times Structure)_i + \varepsilon_i. \quad (4)\end{aligned}$$

The main advantage of HLM is its treatment of error variance: Whereas equation 4 contains only one error term, equations 1, 2, and 3 partition error variance into within-school and between-school components. OLS confounds the two sources of error, a problem that is particularly serious when one level of observations is clustered within a second, as when students are surveyed within schools. This violates the assumption of independent errors in the individual-level model (equation 4), leading to underestimated standard errors (Goldberger and Cain 1982). By estimating separate school-level and student-level errors, HLM adjusts for the correlation of errors within schools (Bryk and Raudenbush 1992).

Another benefit of using HLM is that it estimates the total school-level variance in β_{0j} and β_{1j}, before and after the multilevel interactions are included. I use this feature to indicate the degree to which sector and structural variables account for net between-school variation in achievement and track effects. I also use it to test for the heterogeneity of achievement means and distributions across schools.

Of course, HLM does not resolve every statistical difficulty in estimating the effects of schooling. One especially important issue for this study not specifically addressed by HLM is that students are assigned to tracks on the basis of anticipated differences in the very outcomes in which we are interested. If this differential selection to tracks is not taken into account, then what appear to be track effects may simply reflect pre-existing differences among students enrolled in the different tracks. A statistically related problem is that controls for prior conditions are not com-

---

[5] In equation 1, β_{2j} represents a number of background variables; I have written the equation as if there were only one for the sake of simplicity. In principle, the effects of the background variables could also be allowed to vary between schools. However, freeing more slopes multiplies the number of variances and covariances that are estimated, dramatically increasing the complexity of the model and the difficulty of estimation. For this reason, HLM users are advised to start small, freeing parameters only when there is theoretical interest in their variability (Bryk and Raudenbush 1992).

[6] Again, for simplicity I have written the equations as if there were but one structural predictor, although several will be included in the analyses.

[7] To adjust the within-school intercepts for variation in effects permitted to vary across schools, the within-school variable must be centered around each school's mean and the school's average for the variable must be included in the equation for the intercept (equation 2) (Bryk and Raudenbush 1992). In this analysis, background variables have constant effects across schools, so they are centered around their grand means, i.e., they are deviated from the means of the total sample. Because the effects of tracking are allowed to vary across schools, the track variable is centered around school means. Later in the analysis, the proportion of students in the academic track is included as a predictor of adjusted school achievement.

VARIABLE EFFECTS OF HIGH SCHOOL TRACKING

pletely reliable, which reduces their effectiveness and possibly inflates the estimates of track effects. My approach to dealing with selection bias and unreliability in this analysis is to include a rich set of controls for observed prior conditions in the student-level equation. Although no set of controls can eliminate selection bias with certainty, previous research using these same data has indicated that a comparable array of control variables eliminates the correlation between unobserved selection factors and subsequent achievement, suggesting that selection bias can be effectively reduced in this case (Gamoran and Mare 1989). Jencks (1985) also advocated inclusion of multiple prior test scores to compensate for unreliability in these data.

Analyses were conducted using the HLM computer program (Bryk, Raudenbush, Seltzer, and Congdon 1988). Kreft, Kim, and DeLeeuw (1990) provided a comparison of HLM with other programs for multilevel analysis. I begin by examining the extent to which schools vary in net average achievement and in the net effect of membership in the academic program on achievement. I then explore the sectoral and structural sources of between-school variation in these parameters.

DATA

For information on tracking, achievement, and school characteristics, the best data set available is High School and Beyond (HSB), a survey of a national sample of high schools and students begun in 1980 (Jones et al. 1983). For the present analyses, I use data from 964 public and Catholic schools in the 1980 (base year) and 1982 (first follow-up) surveys. Data were gathered from a random sample of up to 36 students in each school, for a total of 28,804 students. I deleted 11 schools that had 10 or fewer student respondents. I also eliminated 30 schools that had no college-track students and 4 schools in which all students reported belonging to the college track. By drawing on information from other groups, HLM can estimate parameters for a variable that has no within-group variance (e.g., when all students are in the same track). However, the study concerns the impact of differences in the structure of tracking, not the presence or absence of tracking. Moreover, some of the structural variables were undefined when all students reported the same track, resulting in school-level missing data. Additional missing data at the school level reduced the sample to 883 schools (805 public and 78 Catholic), or 91.6 percent of the original sample. All students with data on track position and 1982 achievement were included in the student-level sample, for a total of 20,762, or 77.3 percent of the students surveyed in the 883 schools. Missing values on student-level independent variables were imputed using regressions based on the variables with data present.

Achievement Outcomes

Senior-year (1982) scores on multiple-choice tests of mathematics and verbal achievement serve as two separate individual-level outcomes. Heyns and Hilton (1982) reported reliabilities of .85 (part I) and .54 (part II) for the 38-item, two-part math test. I summed the two parts to create a single measure of math achievement. Verbal achievement was calculated by adding scores on the 20-item reading test and the 21-item vocabulary test, for which Heyns and Hilton (1982) reported reliabilities of .78 and .81, respectively. Table 1 provides sample means and standard deviations for all variables.

Track Positions

Students' track positions are indicated by their self-reported membership in an academic or nonacademic program in their sophomore year (1980). Use of sophomore-year track reports eliminates the problem of whether senior-year reports, which are also available in the data, are a response to achievement rather than a cause. Student reports do not always agree with school officials' reports of track locations, presumably because many schools do not formally label their tracks (Moore and Davenport 1988) and students are not always aware of their overall curricular programs (Rosenbaum 1980). However, prior research demonstrated that self-reports are relevant for a study of track effects on achievement. Self-reports are likely to capture the social-psychological aspects of tracking because track perceptions are linked to expectations and peer associations (Gamoran and Berends 1987; Hallinan and Williams 1989). Self-reports may be less sensitive to instructional differences associated with tracking, but previous work showed they correspond reasonably well to courses taken. A 1972 national survey found that students' and school officials' reports of track positions agreed in about 80 percent of cases (Fennessey, Alexander, Riordan, and Salganik 1981). Vanfossen, Jones, and Spade's (1987) analysis of data from a decade later indicated that 85 percent of students

Table 1. Means and Standard Deviations of Variables in the Analysis: High School and Beyond Survey, 1980 and 1982

Variable	Mean	S.D.
Student-level variables		
Math achievement, 1982	20.066	8.099
Verbal achievement, 1982	22.772	8.101
Math achievement, 1980	18.896	7.121
Verbal achievement, 1980	20.222	7.299
Science achievement, 1980	11.095	3.577
Writing achievement, 1980	10.417	3.783
Sex (1 = female)	.512	.500
Ethnicity (1 = black or Hispanic)	.233	.423
Socioeconomic status	−.050	.709
Academic track	.336	.472
School-level variables		
Catholic	.099	.299
School mean socioeconomic status	−.116	.367
Selectivity		
Achievement gap (math)	5.017	3.925
Achievement gap (verbal)	5.129	4.116
Track heterogeneity (math) (log)	1.536	.174
Track heterogeneity (verbal) (log)	1.580	.147
Electivity		
Proportion choosing own track	.644	.185
Inclusiveness		
Proportion in academic track	.307	.208
(Proportion in academic track)^2	.137	.179
Scope		
Track immobility	.437	.264
Honors rigidity	.443	.247
Remedial rigidity	.528	.221

Note: Means and standard deviations were computed using High School and Beyond design weights (Jones et al. 1983). Unweighted observations are 20,762 students and 883 schools.

who reported the college track as sophomores took math and science courses that were possibly or definitely college-oriented. By contrast, 64 percent of nonacademic-track students took math and science courses that were definitely not college-oriented. Similarly, Gamoran (1987) showed that students who said they were in the college track took more academic courses and more advanced academic courses, especially in math and science, compared with other students.

Other Student-Level Variables

The student-level equation (equation 1) describes the predictors of achievement within each school. In addition to track position, it includes three items drawn from student questionnaires: sex, minority status (black or Hispanic), and socioeconomic status (a composite consisting of the mean of nonmissing standardized values for mother's and father's education, father's occupation, family income, and home artifacts). Equation 1 also includes sophomore-year performance on the math and verbal tests, as well as on tests of science and writing achievement. These control variables are associated with tracking and with achievement, and they are included to purge the estimated track differences from differences in the types of students assigned to various tracks. Using a similar set of within-school predictors of track locations, Gamoran and Mare (1989) found that estimates of bias due to differential selection to tracks were reduced to nearly zero.

School-Level Variables

Catholic schools are indicated by a dummy variable coded 1 (versus 0 for public schools). I also calculated a measure of school mean socioeconomic status by averaging student socioeconomic status within schools. This variable is included as a control variable when estimating effects on school mean achievement (productivity), so that apparent effects of Catholic schools and structural conditions do not simply reflect differences in the socioeconomic contexts of the schools. In equation 2, school mean socioeconomic status is a "contextual effect," i.e., an effect of school socioeconomic status on average achievement in the school over and above the effect of individual socioeconomic status on student achievement within the school (Heyns 1986). The student-level control variables adjust mean achievement for "compositional" differences, and school mean socioeconomic status is included so that sectoral and structural effects are estimated apart from "contextual" effects, which may operate through mechanisms not addressed in this study. The remaining school-level variables describe the structural dimensions of tracking.

Selectivity. I constructed two types of selectivity indicators. One type reflects the initial achievement gaps between tracks, computed for each school as the difference between the average test scores of college-track sophomores and those of noncollege-track sophomores. The second type, track hetereogeneity, is the pooled within-track variance in sophomore test scores for each school. Because of a high negative skew, I transformed the variances logarithmically. I computed two

VARIABLE EFFECTS OF HIGH SCHOOL TRACKING

indicators for each type, one for math achievement and one for verbal achievement. Large achievement gaps between tracks and less within-track heterogeneity (i.e., smaller track variances) indicate more selective tracking systems.

On average, the sophomore achievement gap between tracks was about 5 points in both subjects, but the average pooled within-track variance (about 39) was almost as large as the typical total school variance (about 45). This suggests a possible weakness in the selectivity measures; there may be further differentiation within tracks that these measures do not capture (Oakes 1985).

Electivity. Electivity is the proportion of students in the school who said they chose their curricular programs. This measure relies on students' perceptions of electivity, but as I noted earlier, students' perceptions of whether they chose their tracks are probably more relevant for the tracking-achievement relation than the objective circumstances through which assignment occurs. Table 1 shows that, on average, nearly two-thirds of students in a school believe they chose their tracks.

Inclusiveness. Inclusiveness is indicated by the proportion of students in a school's academic track.[8] To allow for the anticipated nonlinear effects of inclusiveness, I included a quadratic term for this variable. Inequality between tracks is expected to be greatest when inclusiveness is very high or very low; this would be indicated by a negative linear effect and a positive quadratic term. School productivity is expected to rise with increasing inclusiveness, but at a declining rate; this would be indicated by a positive linear effect and a negative quadratic term.

Scope. I calculated three indicators of scope. The first, track immobility, is a measure of agreement between students' sophomore-year and senior-year track positions. This variable is a kappa statistic (Cohen 1960) — it indicates the extent to which students tend to remain in the same track over time.[9] I use the kappa statistic rather than simple proportions of students moving in and out of the college track because it is independent of differences in the marginal distributions of students across tracks. A kappa value of 1 indicates no mobility between the sophomore and senior years, whereas 0 indicates students were as likely to move as to stay. (Negative values for kappa are also possible, but they are unlikely in this situation because they would indicate a tendency for students to shift tracks more often than remaining). The kappa statistic for track immobility was computed separately for each school. Table 1 shows an average of .437, indicating a moderate amount of mobility, consistent with previous work (Gamoran 1987).

The other two indicators are also kappa statistics: Honors rigidity is the extent to which students who reported taking honors math classes also take honors English classes; remedial rigidity is the extent to which students in remedial math also take remedial English. High values for track immobility, honors rigidity, and remedial rigidity indicate wide scope in tracking systems.

RESULTS

I estimated three HLM models for each of the two achievement outcomes. Model 1 is a baseline model, which produces estimates for the within-school equation (equation 1), and for the variance components of the parameters that differ among schools (i.e., β_{0j} and β_{1j} from equations 1, 2, and 3).

Model 2 adds sector (Catholic versus public schools) as a predictor of between-school differences in track effects and in mean achievement, adjusted for differences in composition. (Model

---

[8] Kilgore's (1991) measure of inclusiveness was the proportion of students in a school in the academic track adjusted for achievement. In her study, inclusiveness was a dependent variable. Here inclusiveness is an independent variable, and its effects are estimated on outcomes that are adjusted for prior achievement and background variables. Conceptually, the simple proportion is appropriate for my purposes because students' notions of where they are in the academic hierarchy are likely to be influenced by the absolute size of the academic track in their school, not by the size of their academic track relative to that in other schools with similar compositions (Page 1991).

[9] The formula for kappa is $(P_o - P_c)/(1 - P_c)$ where P_o is the proportion observed and P_c is the proportion expected by chance (Agresti 1990). For example, track immobility in a school is computed as:

$$\frac{\left[(P_o Ac + P_o NAc) - (PAc80 \times PAc82 + PNAc80 \times PNAc82)\right]}{\left[1 - (PAc80 \times PAc82 + PNAc80 \times PNAc82)\right]},$$

where $P_o Ac$ is the proportion in the academic track in both years; $P_o NAc$ is the proportion in the nonacademic track in both years; *PAc80* and *PAc82* are the proportions in the academic track in 1980 and 1982; and *PNAc80* and *PNAc82* are the proportions in the nonacademic track in 1980 and 1982. (Multiplying and summing the marginals as indicated yields the cell proportions expected by chance.)

Table 2. Gamma Coefficients from HLM Analyses of 1982 Math Achievement and 1982 Verbal Achievement: High School and Beyond Survey, 1980 and 1982

	1982 Math Achievement			1982 Verbal Achievement		
Predictor Variable	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
STUDENT-LEVEL EQUATION						
Math achievement, 1980	.666*** (.007)	.663*** (.007)	.660*** (.007)	.095*** (.007)	.091*** (.007)	.090*** (.007)
Verbal achievement, 1980	.089*** (.007)	.085*** (.007)	.082*** (.007)	.589*** (.007)	.585*** (.007)	.584*** (.007)
Science achievement, 1980	.153*** (.014)	.155*** (.014)	.158*** (.014)	.246*** (.013)	.250*** (.013)	.250*** (.013)
Writing achievement, 1980	.159*** (.013)	.156*** (.013)	.154*** (.013)	.244*** (.013)	.240*** (.013)	.238*** (.013)
Sex (1 = female)	-.804*** (.070)	-.809*** (.070)	-.817*** (.070)	-.190** (.069)	-.196** (.069)	-.203** (.069)
Ethnicity (1 = black or Hispanic)	-.641*** (.082)	-.581*** (.082)	-.637*** (.083)	-.926*** (.080)	-.899*** (.081)	-.908*** (.082)
Socioeconomic status	.742*** (.050)	.600*** (.053)	.582*** (.053)	.708*** (.050)	.573*** (.052)	.569*** (.052)
Academic track	1.440*** (.080)	1.592*** (.085)	.047 (.731)	.940*** (.076)	1.047*** (.081)	.139 (.864)
Intercept (adjusted school achievement)	20.133*** (.044)	20.116*** (.046)	19.069*** (.446)	22.748*** (.043)	22.692*** (.043)	21.815*** (.469)
SCHOOL-LEVEL EQUATIONS						
Effects on between-track inequality						
Catholic	—	-.665** (.245)	-.372 (.286)	—	-.284 (.232)	.102 (.287)
Selectivity						
Achievement gap	—	—	-.026 (.023)	—	—	-.032 (.022)
Track heterogeneity	—	—	.066 (.404)	—	—	.019 (.479)
Electivity						
Proportion choosing own track	—	—	.804 (.490)	—	—	.041 (.483)
Inclusiveness						
Proportion in academic track	—	—	-3.235 (1.737)	—	—	.180 (1.734)
(Proportion in academic track)^2	—	—	4.830* (1.974)	—	—	-.901 (1.970)
Scope						
Track immobility	—	—	2.811*** (.374)	—	—	1.625*** (.381)
Honors rigidity	—	—	.262 (.315)	—	—	.387 (.316)
Remedial rigidity	—	—	-.084 (.330)	—	—	.216 (.330)
Effects on adjusted school achievement (productivity)						
Catholic	—	.839*** (.145)	.530** (.177)	—	1.119*** (.137)	1.064*** (.167)
School mean socioeconomic status	—	.917*** (.125)	.614*** (.144)	—	.770*** (.119)	.542*** (.132)
Selectivity						
Achievement gap	—	—	.007 (.012)	—	—	-.013 (.011)
Track heterogeneity	—	—	.358 (.247)	—	—	.262 (.259)
Electivity						
Proportion choosing own track	—	—	-.380 (.279)	—	—	.508 (.262)
Inclusiveness						
Proportion in academic track	—	—	3.784*** (.823)	—	—	1.600* (.777)
(Proportion in academic track)^2	—	—	-2.852** (.916)	—	—	-.916 (.865)
Scope						
Track immobility	—	—	-.520* (.210)	—	—	.040 (.203)
Honors rigidity	—	—	-.184 (.180)	—	—	-.311 (.171)
Remedial rigidity	—	—	.355 (.191)	—	—	-.208 (.181)

$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

Note: Numbers in parentheses are standard errors; N = 883 schools and 20,762 students.

2 also includes school mean socioeconomic status as a predictor of school achievement.) In Model 3, structural dimensions of tracking are added to show how these organizational conditions affect school achievement and the gaps between tracks in achievement. Table 2 displays the results of the three models for the each of the two achievement outcomes (1982 math scores and 1982 verbal scores).

Baseline Models

The baseline model (in Table 2 Model 1) presents estimates for the student-level equation. Each of the eight predictors exerts a significant impact on the math and verbal achievement scores.[10] Consistent with previous research, the average effect of academic-track membership is positive, reflecting increasing between-track inequality.

The baseline model also provides estimates of residual variance for the two coefficients that were allowed to vary between schools: the academic track effect, and school mean achievement adjusted for the composition of the student body. These parameters are displayed in the top panel (Model 1) of Table 3. The chi-square tests indicate significant variation between schools in adjusted mean achievement in both subjects. The impact of tracking on math achievement also varies significantly between schools, but the degree of variation in tracking's effect on verbal achievement is much smaller and is not statistically significant. Thus, additive models of track effects in previous research have been appropriate for examining verbal achievement but incomplete for studying math achievement.

The next step is to try to account for observed variation among schools in adjusted mean achievement and track effects. Although there is little variation to explain in the case of track effects on verbal achievement, I assess the model for that parameter as well as the others for comparative purposes.

Sector Effects

Model 2 in Table 2 addresses the question of whether track effects and overall achievement vary between Catholic schools and public schools. In the analysis of between-track inequality, the Catholic-school coefficients are negative, indicating smaller achievement gaps between tracks in Catholic schools, but the effect is statistically significant only for math achievement. In that subject, academic-track students in public schools differ from their nonacademic counterparts by 1.592 points, net of background and prior achievement. For students in Catholic schools, the difference between tracks is only (1.592 – .665) = .927 points, or about 42 percent smaller.

In the analysis of adjusted school achievement, the results show the familiar Catholic-school advantage on both math and verbal tests, even after allowing for the positive contextual effect of mean socioeconomic status. Especially in math, then, Catholic schools have less inequality between tracks in a context of higher overall achievement, supporting Hypotheses 5a and 5b. The differences between Catholic schools and public schools may result in part from differences in thestructure of tracking in the two sectors.

Effects of the Organization of Tracking

Model 3 in Table 2 shows the impact of the structural characteristics of tracking on the achievement gaps between tracks and on adjusted school achievement. Track immobility, a measure of scope indicating whether students tend to remain in the same tracks over time, leads to greater inequality between tracks in both math and verbal achievement. Other things being equal, the gap between tracks in a very rigid tracking system (defined as one standard deviation above the mean, or a kappa statistic of .701) is wider than the gap between tracks in a very flexible system (one standard deviation below the mean, or kappa = .173) by almost 1.5 points in math and more than 0.8 points on the verbal test. In math, track immobility also reduces achievement overall, but this finding is not replicated for verbal achievement. Thus, Hypothesis 4a is supported for both subjects and Hypothesis 4b is supported only for math.

Between-track inequality in math achievement is greater when inclusiveness is high or low, and smaller when inclusiveness is moderate. This is indicated by the negative linear coefficient (–3.235) and positive quadratic coefficient (4.830) for the proportion of students in the academic track. This finding supports Hypothesis 3a. What do these coefficients mean in substantive terms? Evaluating the effects of inclusiveness at the sample averages for all other school-level variables yields

---

[10] The student-level results agree with prior studies except for the negative effect of being female on verbal achievement. Previous analyses of these data found no significant sex differences in reading and vocabulary scores (Gamoran 1987). The negative sex effect reflects the inclusion of prior writing achievement, an area in which females have a substantial advantage.

Table 3 Chi-Square Tests for Homogeneity of Parameter Variance in the HLM Models: High School and Beyond Survey 1980 and 1982

Model and Parameter	Residual Variance	D.F.	χ^2	Percent of Variance Explained
MODEL 1 (BASELINE EFFECTS)				
Math				
Track effect	551	872	973.4**	—
Adjusted school achievement	.881	872	1819.4***	
Verbal				
Track effect	243	872	903.3	—
Adjusted school achievement	.757	872	1700.9***	
MODEL 2 (SECTOR EFFECTS)				
Math				
Track effect	591	871	969.9*	0.0
Adjusted school achievement	.711	870	1631.8***	19.3
Verbal				
Track effect	263	871	902.7	0.0
Adjusted school achievement	.561	870	1479.5***	25.9
MODEL 3 (ORGANIZATIONAL EFFECTS)				
Math				
Track effect	107	863	880.6	80.6
Adjusted school achievement	.664	862	1562.7***	24.7
Verbal				
Track effect	207	863	877.0	15.1
Adjusted school achievement	.532	862	1436.3***	29.8

$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

the following results: When inclusiveness is at the sample mean (.307), the gap between the college track and the noncollege track in math achievement is 1.26 points; when inclusiveness is low (.10), the between-track gap increases to 1.52 points; but when inclusiveness is very high (.75), between-track inequality also increases to as much as 2.09 points on the math test. This pattern, however, does not hold for verbal achievement.

Inclusiveness of a tracking system also affects school mean achievement, not only in math but on the verbal test as well. The positive linear effect and negative quadratic coefficient are consistent with a positive impact at a declining rate as specified by Hypothesis 3b. Thus, Hypothesis 3b is supported for both subjects and Hypothesis 3a is supported for math.

In contrast to the hypotheses for sector, scope, and inclusiveness, which are generally supported, I found no support for hypotheses about the electivity and selectivity of tracking systems. Most coefficients for electivity and selectivity are small, and none are statistically significant. Hence, the data are not consistent with Hypotheses 1a, 1b, 2a, or 2b.

For math achievement, the Catholic-school effects on the impact of tracking and on adjusted school achievement decline from Model 2 to Model 3 (as does the effect of school socioeconomic status on mean achievement). The sector difference in inequality drops by 44 percent (–.665 to –.372) and is no longer statistically significant, and the Catholic-school advantage in mean achievement drops by 37 percent (.839 to .530). This pattern is consistent with the argument that lower inequality and higher productivity of Catholic schools result in part from differences in the structure of tracking. The pattern, however, is weakly replicated in the verbal analysis.

Given the much greater variation in between-track inequality for math achievement as compared to verbal achievement, it is not surprising that I had greater success explaining between-school variation in the track effect on math achievement. In Table 3, the bottom panel (Model 3) shows the decline in residual variance after the school-level predictors are added. About 80 percent of between-school variation in track effects on math achievement is explained, compared to only around 15 percent of the variance in track effects on verbal achievement. With sector and structural conditions taken into account, remaining variation in the track effect on math achievement is no longer statistically significant. For adjusted school achievement, the final model accounts for about 25 percent of the variance in math scores and nearly 30 percent of the variance in mean verbal achievement, and significant residual variation remains.

DISCUSSION AND CONCLUSIONS

In this study, the question of how tracking affects achievement elicits a more complex answer than it has in the past. In general, the analyses indicate that the effects of tracking depend in part on the structure of the tracking system. This claim is best supported for math achievement, but it also holds to some extent for verbal achievement. Schools with less mobility in their tracking systems tend to have greater between-track inequality in both subjects, and they have lower overall math scores. Tracking systems that are high or low in inclusiveness also exhibit wider gaps between tracks in math achievement. Average math and verbal scores are higher in more inclusive systems, although the gains from inclusiveness accrue at a declining rate. Finally, Catholic schools not only have higher overall achievement, net of measured background variables, but for math they also have less inequality between tracks, supporting previous speculation (Gamoran and Berends 1987).

Why are the patterns generally sharper for math achievement? Although track *effects* on inequality are evident for both verbal and math achievement, the *variability* of track effects is insignificant for the verbal test. I posited two mechanisms for track effects, reflecting social-psychological and instructional differentiation. Between-school differences in social-psychological mechanisms should be no less salient for verbal achievement than for math achievement. However, instructional differentiation may be more variable among schools in math than in English. Hence, the more variable effects of tracking on math achievement may reflect greater between-school differences in the organization of math instruction. Still, track immobility contributed to increased between-track inequality in verbal as well as math achievement. This finding is consistent with the view that peer-group effects, which may be accentuated by permanent track assignments, are linked to inequality between tracks in both subjects.

Although the residual variance was initially greater for track effects on math achievement, the variance that remained unexplained was greater for verbal achievement. Other aspects of tracking systems, unexamined in this study, may explain varied track effects on verbal achievement. For example, schools may vary in the culture or ethos of tracking — tracking may be a clear symbol of students' future directions in some schools, while its significance in other schools is more vague (Gamoran and Berends 1987). Presumably, tracking's impact would be magnified where its power to confer status is greater. Symbolic differences among schools, which were not addressed in the data, may be linked to variation in track effects on verbal achievement (Lamont and Lareau 1988).

Some structural characteristics of tracking systems did not exhibit the predicted effects. The inconsistent and insignificant effects of electivity may indicate that two processes cancel out: Greater electivity may lead to increased motivation, as I argued, but may also promote between-track differences in aspirations as specified by Sørensen (1970). This issue could be addressed by examining peer-group formation in elective and nonelective tracking systems. Such analysis would show whether elective systems promote more homogeneous friendship groups that in turn lead to more powerful peer-group effects, as Sørensen predicted. The study of peer groups in different types of tracking systems may also reveal whether more permanent track assignments encourage within-track friendship formation, as Sørensen also argued, a prediction that is consistent with my results for track immobility.

The absence of effects for selectivity may reflect a weakness of the measures, particularly the indicator of track heterogeneity: The conceptual model refers to heterogeneity of students' *classes*, but the data address only the heterogeneity of tracks. If tracks are more finely differentiated than my measures reveal — Oakes (1985), for example, described ability-grouping within tracks as common — then the analysis may have missed the actual impact of reducing heterogeneity for track effects and for average achievement.

Overall, my findings underscore the importance of assessing contextual variation in microsocial processes. Although the results are consistent with prior research for the general case — on average, belonging to the academic track is beneficial for achievement — the advantage is not the same in all schools, at least not for math achievement. The academic-track advantage is less in schools with more flexible and (for math) moderately inclusive tracking systems. At the same time, my results should not encourage a haphazard search for contextual variation. An a priori conceptual framework should suggest what dimensions of the context need to be examined. In this study, the framework for understanding the aggregate-level differences was built on knowledge of how the microlevel processes occur.

The results also draw attention to the value of examining tracking's variable effects on productivity as well as inequality. The finding of less between-track inequality in math scores in Catholic schools, for example, does not by itself indicate that Catholic schools used tracking more successfully. The results could have occurred through lower scores in the academic track. By studying productivity as well, I confirmed that lower inequality occurred along with higher overall achievement, suggesting that the narrower gap in Catholic schools occurs because low-track students are brought up, not because high-track students are held down (Hoffer et al. 1985). Conversely, the results for track immobility suggest that inflexible tracking systems have greater inequality along with lower average achievement, presumably reflecting especially poor conditions in the nonacademic tracks of such schools. The implications of the findings for inclusiveness are even more complex: In math, between-track inequality is lower when inclusiveness is moderate, but productivity is higher when inclusiveness is high. Consequently, an educator must choose between maximizing overall achievement in the school — usually a significant goal — and minimizing inequality between tracks within a school.

Quantitative analysis makes the world appear simpler than it really is. Are the benefits of added complexity worth the difficulties of conceptualization, estimation, and interpretation? In this case, I think the enhanced theoretical understanding and the potential policy benefits justify the effort, particularly with regard to math achievement.

ADAM GAMORAN *is Professor of Sociology and Educational Policy Studies at the University of Wisconsin, Madison. His main research interest has been the effects of stratification in school systems, especially the relation between tracking and achievement, and the role of classroom instruction as a mechanism underlying differences between tracks in achievement. He is spending the 1992–1993 academic year as a Fulbright Scholar at the University of Edinburgh, where he is studying the impact of curriculum standardization on changes in levels and inequality of attainment in Scottish secondary education.*

REFERENCES

Agresti, Alan. 1990. *Categorical Data Analysis*. New York: Wiley.

Alexander, Karl L. and Aaron Pallas. 1985. "School Sector and Cognitive Performance: When Is a Little a Little?" *Sociology of Education* 58:115–27.

Alexander, Karl L., and Martha A. Cook. 1982. "Curricula and Coursework: A Surprise Ending to a Familiar Story." *American Sociological Review* 47:626–40.

Ball, Stephen J. 1981. *Beachside Comprehensive: A Case-Study of Secondary Schooling*. Cambridge, England: Cambridge University Press.

Barr, Rebecca and Robert Dreeben. 1983. *How Schools Work*. Chicago: University of Chicago Press.

Berends, Mark. 1991. *High School Tracking and Students' School Orientations*. Department of Sociology, University of Wisconsin, Madison. Unpublished manuscript.

Bryk, Anthony S. and Mary E. Driscoll. 1988. *The High School as Community: Contextual Influences, and Consequences for Students and Teachers*. Madison, WI: National Center on Effective Secondary Schools.

Bryk, Anthony S., and Stephen W. Raudenbush. 1992. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA: Sage.

Bryk, Anthony S., Stephen W. Raudenbush, Michael Seltzer, and Richard T. Congdon. 1988. *An Introduction to HLM: Computer Program and Users' Guide. Version 2.2*. Chicago: Department of Education, University of Chicago.

Burgess, Robert G. 1983. *Experiencing Comprehensive Education: A Study of Bishop McGregor School*. London: Methuen and Co.

_______. 1984. "It's Not a Proper Subject: It's Just Newsom." Pp. 181–200 in *Defining the Curriculum*, edited by I. Goodson and S. Ball. London: The Falmer Press.

Camarena, Margaret. 1990. "Following the Right Track: A Comparison of Tracking Practices in Public and Catholic Schools." Pp. 159–82 in *Curriculum Differentiation: Interpretive Studies in U. S. Secondary Schools*, edited by R. Page and L. Valli. Albany, NY: SUNY Press.

Cicourel, Aaron V. and John I. Kitsuse. 1963. *The Educational Decision-Makers*. Indianapolis, IN: The Bobbs Merrill Co.

Cohen, Jacob. 1960. "A Coefficient of Agreement for Nominal Scales." *Educational and Psychological Measurement* 20:37–46.

Dahloff, Urban. 1971. *Ability Grouping, Content Validity, and Curriculum Process*. New York: Teachers College Press.

DiPrete, Thomas A. and David B. Grusky. 1991. "Structure and Trend in the Process of Stratification for American Men and Women." *American Journal of Sociology* 96:107–43.

Eckert, Penelope. 1989. *Jocks and Burnouts: Social Categories and Identity in the High School*. New York: Teachers College Press.

Fennessey, James, Karl L. Alexander, Cornelius Riordan, and Laura Hersh Salganik. 1981. "Tracking and Frustration Reconsidered: Appearance or Reality." *Sociology of Education* 54:302–09.

Finley, Merilee K. 1984. "Teachers and Tracking in a Comprehensive High School." *Sociology of Education* 57:233–43.

Gamoran, Adam. 1984. "Teaching, Grouping, and Learning: A Study of the Consequences of Educational Stratification." Ph. D. dissertation, Department of Education, University of Chicago, Chicago, IL.

_______. 1987. "The Stratification of High School Learning Opportunities." *Sociology of Education* 60:135–55.

VARIABLE EFFECTS OF HIGH SCHOOL TRACKING

________. 1992. "Access to Excellence: Assignment to Honors English Classes in the Transition from Middle to High School." *Educational Evaluation and Policy Analysis* 14:185–204.

Gamoran, Adam and Mark Berends. 1987. "The Effects of Stratification in Secondary Schools: Synthesis of Survey and Ethnographic Research." *Review of Educational Research* 57:415–35.

Gamoran, Adam, Martin Nystrand, Mark Berends, and Paul C. LePore. 1992. "Beyond Technical Rationality: Ability Grouping and the Distribution and Effects of Classroom Instruction" (Discussion Paper No. 981–92). Institute for Research on Poverty, Madison, WI.

Gamoran, Adam and Robert D. Mare. 1989. "Secondary School Tracking and Educational Inequality: Reinforcement, Compensation, or Neutrality?" *American Journal of Sociology* 94:1146–83.

Goldberger, Arthur S. and Glen G. Cain. 1982. "The Causal Analysis of Cognitive Outcomes in the Coleman, Hoffer, and Kilgore Report." *Sociology of Education* 55:103–22.

Hallinan, Maureen T. and Richard A. Williams. 1989. "Interracial Friendship Choices in Secondary Schools." *American Sociological Review* 54:67–78.

Hargreaves, David H. 1967. *Social Relations in a Secondary School*. London: Tinling.

Hauser, Robert M., William H. Sewell, and Duane F. Alwin. 1976. "High School Effects on Achievement. Pp. 309–41 in *Schooling and Achievement in American Society*, edited by W. H. Sewell, R. M. Hauser, and D. L. Featherman. New York: Academic Press.

Heyns, Barbara. 1974. "Social Selection and Stratification within Schools." *American Journal of Sociology* 79:1434–51.

________. 1986. "Educational Effects: Issues in Conceptualization and Measurement." Pp. 305–40 in *Handbook of Theory and Research for the Sociology of Education*, edited by J. G. Richardson. New York: Greenwood.

Heyns, Barbara and Thomas L. Hilton. 1982. "The Cognitive Tests for High School and Beyond: An Assessment." *Sociology of Education* 55:89–102.

Hoffer, Thomas, Andrew M. Greeley, and James S. Coleman. 1985. "Achievement Growth in Public and Catholic Schools." *Sociology of Education* 58:74–97.

Jencks, Christopher L. 1985. "How Much Do High School Students Learn?" *Sociology of Education* 58:128–35.

Jencks, Christopher, Marshall Smith, Henry Acland, Mary Jo Bane, David Cohen, Herbert Gintis, Barbara Heyns, and Stephan Michaelson. 1972. *Inequality: A Reassessment of the Effects of Family and Schooling in America*. New York: Basic Books.

Jones, Calvin, M. Clark, G. Mooney, H. McWilliams, I. Crawford, B. Stephenson, and R. Tourangeaou. 1983. *High School and Beyond 1980 Sophomore Cohort First Follow-up (1982) Data File User's Manual*. Washington, DC: National Center for Education Statistics.

Jones, James D., Beth E. Vanfossen, and Joan Z. Spade. 1986. "Curriculum Placement: Individual and School Effects Using the High School and Beyond Data." Paper presented at the annual meeting of the American Sociological Association, August, 1986, Washington, DC.

Keddie, Nell. 1971. "Classroom Knowledge." Pp. 133–60 in *Knowledge and Control*, edited by M. F. D. Young. London: Collier-Macmillan.

Kerckhoff, Alan C. 1986. "Effects of Ability Grouping in British Secondary Schools." *American Sociological Review* 51:842–58.

Kilgore, Sally B. 1991. "The Organizational Context of Tracking in Schools." *American Sociological Review* 56:189–203.

Kreft, Ita G. G., K. Kim, and J. DeLeeuw. 1990. *Comparing Four Different Statistical Packages for Hierarchical Linear Regression: GENMOD, HLM, ML2, and VARCL*. Los Angeles: Departments of Psychology and Mathematics Social Statistics Program, University of California–Los Angeles.

Lacey, Colin. 1970. *Hightown Grammar*. Manchester: Manchester University Press.

Lamont, Michele and Annette Lareau. 1988. "Cultural Capital: Allusions, Gaps, and Glissandos in Recent Theoretical Developments." *Sociological Theory* 6:153–68.

Lee, Valerie E. and Anthony S. Bryk. 1988. "Curriculum Tracking as Mediating the Social Distribution of High School Achievement." *Sociology of Education* 62:78–94.

________. 1989. "A Multilevel Model of the Social Distribution of High School Achievement." *Sociology of Education* 62:172–92.

Lundgren, Ulf P. 1972. *Frame Factors and the Teaching Process*. Stockholm: Almqvist and Wiksell.

Mason, William M., George Y. Wong, and Barbara Entwistle. 1983. "Contextual Analysis Through the Multilevel Linear Model." Pp. 72–103 in *Sociological Methodology, 1983–84*, edited by S. Leinhardt. San Francisco: Jossey-Bass.

Metz, Mary H. 1978. *Classrooms and Corridors: The Crisis of Authority in Desegregated Secondary Schools*. Berkeley, CA: University of California Press.

Moore, Don and Suzanne Davenport. 1988. *The New Improved Sorting Machine*. Madison, WI: National Center on Effective Secondary Schools.

Natriello, Gary, Aaron Pallas, and Karl L. Alexander. 1989. "On the Right Track? Curriculum and Academic Achievement." *Sociology of Education* 62:109–18.

Oakes, Jeannie. 1985. *Keeping Track: How Schools Structure Inequality*. New Haven, CT: Yale University Press.

Oakes, Jeannie, Adam Gamoran, and Reba N. Page. 1992. "Curriculum Differentiation: Opportunities, Outcomes, and Meanings." Pp. 570–608 in *Handbook of Research on Curriculum*, edited by P. W. Jackson. New York: Macmillan.

Page, Reba N. 1987. "Lower-Track Classes at a College-Preparatory School: A Caricature of Educational Encounters." Pp. 447–72 in *Interpretive Ethnography of Education: At Home and Abroad*, edited by G. Spindler and L. Spindler. Hillsdale, NJ: Erlbaum.

________. 1991. *Lower Track Classrooms: A Curricular and Cultural Perspective*. New York: Teachers College Press.

Page, Reba N. and Linda Valli, eds. 1990. *Curriculum Differentiation: Interpretive Studies in U. S. Secondary Schools*. Albany, NY: SUNY Press.

Powell, Arthur, Eleanor Farrar, and David K. Cohen. 1985. *The Shopping-Mall High School*. Boston: Houghton-Mifflin.

Rehberg, Richard A. and Evelyn R. Rosenthal. 1978. *Class and Merit in the American High School*. New York: Longman.

Rosenbaum, James E. 1976. *Making Inequality*. New York: Wiley.

________. 1980. "Track Misperceptions and Frustrated College Plans: An Analysis of Track and Track Perceptions in the National Longitudinal Survey." *Sociology of Education* 53:74–88.

________. 1984. "The Social Organization of Instructional Groups." Pp. 53–68 in *The Social Context of Instruction*, edited by P. L. Peterson, L. C. Wilkerson, and M. T. Hallinan. Orlando, FL: Academic Press.

Schwartz, Frances. 1981. "Supporting or Subverting Learning: Peer Group Patterns in Four Tracked Schools." *Anthropology and Education Quarterly* 12:99–121.

Shavit, Yossi and David L. Featherman. 1988. "Schooling, Tracking, and Teenage Intelligence." *Sociology of Education* 61:42–51.

Slavin, Robert E. 1987. "Ability Grouping and Achievement in Elementary Schools: A Best-Evidence Synthesis." *Review of Educational Research* 57:293–336.

________. 1990. "Achievement Effects of Ability Grouping in Secondary Schools: A Best-Evidence Synthesis." *Review of Educational Research* 60:471–99.

Sørensen, Åage B. 1970. "Organizational Differentiation of Students and Educational Opportunity." *Sociology of Education* 43:355–76.

Valli, Linda. 1990. "A Curriculum of Effort: Tracking Students in a Catholic High School." Pp. 45–65 in *Curriculum Differentiation: Interpretive Studies in U. S. Secondary Schools*, edited by R. N. Page and L. Valli. Albany, NY: SUNY Press.

Vanfossen, Beth E., James D. Jones, and Joan Z. Spade. 1987. "Curriculum Tracking and Status Maintenance." *Sociology of Education* 60:104–22.

Waitrowski, Michael D., Stephen Hansell, Charles R. Massey, and David L. Wilson. 1982. "Curriculum Tracking and Delinquency." *American Sociological Review* 47:151–60.

Willms, J. Douglas. 1985. "Catholic-School Effects on Academic Achievement: New Evidence from the High School and Beyond Follow-Up Study." *Sociology of Education* 58:98–114.

Wilson, Barry J. and Donald W. Schmits. 1978. "What's New in Ability Grouping?" *Phi Delta Kappan* 59:535–36.

Review of Educational Research
Fall 1991, Vol. 61, No. 3, pp. 379–406

Community Involvement and Disadvantaged Students: A Review

Saundra Murray Nettles
Johns Hopkins University

This review discusses the effects of community involvement on students who face multiple impediments to success in schools. The first part of the article conceptualizes community involvement as a typology of four processes of social change: conversion, mobilization, allocation of resources, and instruction. Illustrations of these processes are drawn from research and programmatic literature. The second part of the article considers the effects of the varied forms of involvement in a review of 13 evaluations of interventions implemented with significant input from community entities. Overall, the studies indicate that programs can have positive effects on school-related behavior and achievement as well as attitudes and risk-taking behavior. The concluding section identifies gaps in the research and offers a framework for future studies.

Communities have always played important roles in students' intellectual and psychosocial development, but in the last decade educators, youth advocates, and policymakers have called for increased community participation to solve the problems of educationally disadvantaged students. Numerous projects are underway, their existence heralded in the popular and professional literature.

However, optimism and involvement have not been matched by systematic efforts to understand these initiatives in the context of evidence about the community's impact on students. This gap in knowledge can be attributed to the isolation of disciplines, the focus on specific projects rather than general components, and the ambiguity of concepts about community. This article introduces a definition of community involvement, provides a needed synthesis of findings from evaluations of community involvement projects, and offers a conceptual framework for future research.

In this article, the term *educationally disadvantaged* is applied to students who face multiple impediments to success in school. Poor African-American and poor Hispanic students comprise the bulk of those considered to be at-risk of negative educational outcomes, such as illiteracy and school dropout.

Natriello, McDill, and Pallas (1990) estimate that in 1988, 25 million of the nation's 63.6 million children under the age of 18 were educationally disadvantaged when any

---

Work on this review was supported by the Center for Research on Effective Schooling for Disadvantaged Students, supported as a national research and development center by funds from the Office of Educational Research and Improvement, U.S. Department of Education. The opinions expressed in this publication do not necessarily reflect the position or policy of the OERI, and no official endorsement should be inferred. Send requests for reprints to the Center for Research on Effective Schooling for Disadvantaged Students, Johns Hopkins University, 3505 North Charles Street, Baltimore, MD 21218. The author gratefully acknowledges Gary Gottfredson, Mary Kolb, and Barbara Wasik, who provided useful suggestions and comments on drafts of this article.

Copyright 1991 by the American Educational Research Association. Reprinted by permission of the publisher.

one of five risk factors (race/ethnicity, poverty, family structure, language background, and mother's education) were used. They project that the number of educationally disadvantaged children will increase substantially by the year 2,020, when the number of impoverished children will be 16.5 million—a 33% increase over the 12.4 million children in poverty in 1987.

Defining Community Involvement

Community involvement consists of the actions that organizations and individuals (e.g., parents, businesses, universities, social service agencies, and the media) take to promote student development. Such community involvement is typically described in terms of specific roles that community actors play in supporting students (see Carnegie Council on Adolescent Development, 1989; Children's Defense Fund, 1986; Constable & Walberg, 1988; Oakes, 1987; W.T. Grant Foundation, 1988). *Community* refers both to locales, such as neighborhoods, and to social interactions (e.g., relations among a network of social service providers), that can occur within or transcend local boundaries.

Nettles (1989) conceptualized these varied forms of involvement as a typology of four change processes: conversion, mobilization, allocation of resources, and instruction. The first, *conversion,* refers to the process of bringing the student from one belief, or behavioral stance, to another. The second process, *mobilization,* includes actions to increase citizen and organizational participation in the educational process. *Allocation* refers to activities wherein community entities provide resources (such as social support and services) to children and youth. Finally, *instruction* embraces actions designed to assist students in their intellectual development or in learning the rules and values that govern social relationships in the community.

In Nettles's formulation, the four processes embrace natural, or unstructured, occurrences of involvement as well as structured actions that constitute projects and formal interventions. Moreover, in interventions, one process may predominate (e.g., as instruction does in tutoring programs) or a combination of two, or more, processes may be evident. Adopting this typology as a framework for the following review of the literature provides a perspective on mobilization, allocation, and instruction. Unfortunately, although the literature is sprinkled with anecdotes about students who suddenly began to achieve or who suddenly ceased to behave in destructive ways as the result of exposure to a powerful message or charismatic person, systematic research on this kind of phenomenon with disadvantaged students is rare. Also absent from the literature are examples of natural occurrences of resource allocation and mobilization. Thus, I will discuss in the following section structured forms of allocation and mobilization as well as formal and informal examples of instruction.

Involvement as Mobilization

Mobilization embraces actions that fall under such labels as citizen participation, neighborhood organizing, partnerships for school reform and improvement, legal action, and social movements. The targets of such involvement are institutions, political jurisdictions, and geographic areas; therefore, effects on students are likely to be indirect. For example, citizen and parental participation on school governing boards may produce changes in the curriculum or in teacher attitudes towards students. These changes, in turn, may affect student achievement levels.

Community Involvement

The predominant focus of literature on mobilization is the improvement of practices that promote change. There are general guides for community action (Alinsky, 1971; Rothman, Erlich, & Teresa, 1976, 1981) and handbooks that suggest highly specific actions to link schools and students with other entities in the community (see for examples, Asche, 1989; Bain & Herman, 1989; Merenda, 1986; Otterbourg, 1986). These guides commonly include principles of practice based on studies of specific cases. One particularly active area of research is citizen participation in school decision making (see Williams, 1989, for a review and synthesis). Recent attention has focused on partnerships between schools and community entities such as businesses, social service agencies, universities, cultural institutions, and community-based organizations.

Partnerships. In 1975, Jesse Jackson initiated and led a national crusade to involve parents, businesses, students, school staff, and other segments of local communities in the pursuit of excellence in education. The crusade eventually led to the PUSH for Excellence (PUSH-EXCEL) Project, a three-year, federally funded demonstration project that established a highly visible network of educational partnerships. The evaluation of the PUSH-EXCEL Project (S.R. Murray, Murray, Gragg, Kumi, & Parham, 1982) documented the extensive grass-roots organizing that preceded the demonstration projects in Chicago, Kansas City, Chattanooga, and Denver. The evaluation also documented the results of the PUSH-EXCEL Project's efforts to develop both a stable, active base of citizen support and a menu of school and community-based activities to produce improvements in student attendance, academic motivation, sense of responsibility, grades, and test scores. The PUSH-EXCEL Project encountered many difficulties (e.g., defining roles of the various partners and establishing mechanisms for the sustained involvement of partners) in its efforts to transform the vision of its founder into concrete applications.

Case studies of other partnerships (e.g., see Levine & Trachtman, 1988; Pine & Keane, 1989) indicate that the problems the PUSH-EXCEL Project experienced are common in school/community alliances. These difficulties can undermine collaboration unless the implementation process includes mechanisms to foster the relationship between partners. In a review of urban school/community alliances, Ascher (1988) cited as critical features in sustaining partnerships:

> commitment, egalitarian decision-making, a sense of ownership by participants at all levels, clarity about roles, clarity and flexibility about both methods and goals, an ability to bridge different institutional cultures, training, and patience concerning the collaborative process itself. (p. 14)

She concluded that the principles of forming and maintaining successful collaboratives are similar across types of partnerships.

Mann (1987a, 1987b) examined business/school partnerships, a popular type of collaboration in the 1980s, in 23 large cities and in a stratified random sample of 85 U.S. public school districts. Data were collected through telephone interviews with superintendents and other officials and through review of documents. Mann found that formal partnerships were concentrated in big cities and were useful in connecting urban schools and their predominantly low-income and minority populations to the business community. However, partnerships competed with other interests (such as local youth organizations) for funds.

Current national mobilizations include the Black Church Project, that is sponsored by the American Association for the Advancement of Science. This project, through a network of 15 sites, trains church staff and volunteers to conduct science and math workshops, public science days, and science and mathematics career days (George, Richardson, Lakes-Matyas, & Blake, 1989). Another major effort is One to One, which has this national goal: "By 1995, every young person who might benefit from a mentoring relationship will have the opportunity to be matched with a caring partner" (One to One, 1990, p. 2). This match is to be accomplished through the formation of local Leadership Councils, pilot neighborhood projects, and the National Mentoring Partnership. Finally, the ASPIRA Association, a community-based organization devoted to improving the status of Hispanic children and youth, has led a number of national efforts, including the organizing of clubs that provide Latino students with opportunities to receive academic and career counseling and to learn leadership skills. Two other national efforts are: the Public Policy Leadership program and the Hispanic Community Mobilization for Dropout Prevention (ASPIRA Association, 1990).

Involvement as Allocation of Resources

Community involvement often entails the allocation of resources to eliminate disadvantages in students' access to resources. For example, court battles to end school segregation were among the first of many community actions to reallocate educational resources. Other actions serve to remove barriers to access, alter the incentive structure, and provide social support for student efforts.

Access to resources. To remove barriers to student use of health and social services, states and cities are placing clinics and other resources in or near public schools. As of Spring 1989, according to a survey conducted by the Center for Population Options (cited in Kirby, Waszak, & Ziegler, 1989), 90 providers were operating 150 health clinics in 32 states and 91 communities. The majority (59%) of students who used the clinics were African American. Twenty-five percent were White, and 12% were Hispanic. The clinics provided a variety of medical, counseling, educational, reproductive, and family planning services and were typically found in low-income areas. Program operators included community health clinics, nonprofit organizations, hospitals, medical schools, departments of public health, and school systems.

Other involvement efforts focus on increasing low-income youths' access to employment and training. An example is the Youth Incentive Entitlement demonstration, that the Manpower Demonstration Research Corporation managed and evaluated from 1978 to 1980. Through the efforts of the Youth Incentive Entitlement demonstration, over 80,000 low-income youths in 17 cities applied for work in jobs paying the minimum wage; in some cities, the employment rates of minority and White youth were equalized. Private businesses accounted for slightly over half of the work sponsors (Walker, 1984).

Incentives for effort and achievement. It is often assumed that the incentive structure for impoverished youth should be altered and that community actors can play a major role in creating and providing incentives, thereby encouraging students to invest in constructive pursuits. Two of the most widely cited examples of community involvement have attempted to provide incentives that will encourage disadvantaged students to graduate from high school and then either attend college or enter the work force.

The first of these is The Boston Compact, that was initiated in 1982 with a formal agreement to the effect that businesses, labor unions, and the Boston city government would provide priority hiring of graduates of the public high schools. In return, the school system contracted to reduce dropout rates, increase attendance, and assure that graduates were competent in basic skills.

The second example is the "I Have A Dream" Foundation, that Eugene Lang established in 1981 with a spontaneous pledge to the graduating sixth-grade class of East Harlem's Public School 121. Lang promised to pay the college costs of each student who finished high school. He subsequently provided the students with various supports to facilitate their efforts to complete school. As of October 1989, the program had been replicated in 32 cities of 23 states, with funds provided by 132 sponsors of 9,000 students (J.M. Sesnick, personal communication, December 6, 1989).

Both of these programs guaranteed valuable long-term incentives for staying in school. In Boston, the incentives were not of sufficient power to reduce the school dropout rates (Hargroves, 1986). Data on the original class of Dreamers from Public School 121 suggested a more positive outcome than that achieved through The Boston Compact. Ninety percent of the students either obtained, or were expecting to obtain, high school or general equivalency diplomas. The expected rate was 75% ("I Have a Dream" Foundation, 1989). However, it is not clear whether the results for the Dreamers were due to the incentives or due to the assistance that the students received from program staff. The next section briefly considers the functions of such support in the lives of disadvantaged children and youth.

Social support. Informal helpers play important roles in many communities. In poor neighborhoods, interpersonal resources may serve as substitutes for, or extensions of, institutional services and supports (McAdoo 1980; Stack, 1974). Despite the apparent strength of naturally occurring support, evidence from a variety of sources suggests that disadvantaged students either have limited access to resourceful adult helpers, rely heavily on peers, have parents and other family members who lack social support, or are impeded by the demands of members of their social networks.

In programs for disadvantaged students, social support can occur informally, in the context of relationships that are structured chiefly to provide academic, psychological, social, or other services. In the evaluation of six school-based clinics, 43% to 63% of the students who used the clinics cited the staff's caring as one of the five most important reasons for using the clinics (Kirby et al., 1989). When support only occurs as a byproduct of another component, it can be unpredictable, episodic, and untargeted. Planned support that provides for sustained, goal-directed relationships is therefore offered in interventions, although the quality of relationships varies. Freedman (1988) examined relationships between elder mentors and at-risk youth in five programs and found three types: primary relationships, in which a high degree of trust, enjoyment, and attachment were apparent; secondary relationships, which exhibited the characteristics of primary relationships, but in a less developed form; and nonsignificant relationships that were marked by distance (see Flaxman, Ascher, & Harrington, 1988, for further discussion of mentoring and other adult/youth relationships).

Involvement as Instruction

The final form of involvement is instruction, which refers to actions that support intellectual development and social learning. Instruction in the community can occur informally as indicated in studies of language socialization and studies of the role of parent as teacher. According to these studies, social interactions in the home and the wider community are important contexts in which children learn emergent literacy skills, self-regulation of cognitive and other tasks, and other skills and behaviors needed for performance in schools and performance in daily communication (Gundlach, Farr, & Cook-Gumperz, 1989; Heath, 1989; Scott-Jones, 1984, 1989).

Instruction in the community can also take place in organized settings, such as tutoring programs, clubs, and teams. According to a recent survey (U.S. Department of Education, 1990), tutoring was the focal component in the estimated 1,701 college-sponsored programs wherein college students tutored or mentored disadvantaged elementary and secondary students. Churches also play a major instructional role through their programs of religious and moral education as well as their structured tutoring programs. For example, in 1986, the Congress of National Black Churches began Project SPIRIT, a pilot program in five African-American churches in Atlanta, Indianapolis, and Oakland. The project was funded by a grant from the Carnegie Foundation and featured academic tutoring, instruction in life skills, and morale building (Carnegie Corporation of New York, 1987–1988).

Other organized community efforts to foster academic performance provide additional resources and support for learning and stimulating students' desire to achieve in school and in other settings. The National Council of la Raza is implementing the Innovative Education Project through Hispanic community-based organizations in Kansas City, Chicago, and Houston (Orum, 1988). The Project is designed to address the academic and nonacademic needs of low-income Hispanic students and their families. The Project on Adolescent Literacy conducted a national search for successful literacy programs for young adolescents (Davidson & Koppenhaver, 1988) and discovered both in-school and summer programs. In addition, it documented programs that community-based organizations were implementing for children disadvantaged by poverty or by limited proficiency in English.

The role of parents and family in both informal and structured instruction has received considerable attention (for reviews, see Epstein, 1987; Epstein & Scott-Jones, 1988; Henderson, 1987; Scott-Jones, 1984; Slaughter & Epps, 1987; Tangri & Moles, 1987; Weisbaum, 1990), and evidence suggests that parental participation in children's efforts to learn in schools as well as in the broader community can have positive effects on students' school achievement. In the context of a general review of the effectiveness of community involvement in improving outcomes for disadvantaged children and youth, the next section includes an illustrative study of parental involvement in a tutoring program.

Review of the Literature

This review discusses the effects of the varied forms of involvement defined above. The studies that are examined were drawn from the empirical literature on interventions that are characterized by significant input from community entities.

Scope of the Review

The studies reviewed below were located through a search of the ERIC and PsychLit databases and through a manual search of current newsletters (e.g., *Education Week*) and journals. All studies identified in the search were examined, if they met the following criteria: (a) if they addressed academic and other effects of programs, or projects, that were developed, or administered, by entities either outside formal educational systems or staffed primarily by community residents or employees of public or private service agencies and (b) if a substantial proportion of the participants in the program came from low-income families or had other characteristics that were associated with educational failure. Studies were selected for review if their designs included some form of comparative analysis and if their reports (published and unpublished) were sufficiently detailed to permit evaluation of sample composition, measures used, level of program implementation, and quality and type of data analysis.

Table 1 summarizes the key components of the interventions and the major features of the research designs. Five of the programs were forms of resource allocation: the Cambridge-Somerville Youth Study (Powers & Witmer, 1951), the Chicago Area Project (Schlossman, Zellman, & Shavelson, 1984), Project Redirection (Polit & Kahn, 1985), the Pregnancy Prevention Program (Zabin, Hirsch, Smith, Streett, & Hardy, 1986), and the Resource Mother Home Visit Program (Unger & Wandersman, 1985). Each of these programs offered several kinds of activities, including counseling or other forms of social support and services, such as information and referral, recreation, or family planning.

Two of the programs employed instruction as the sole or main component. They included EXTRA (Sheley, 1984) and the Parent-Child Tutoring Program (Mehran & White, 1988). Four of the programs combined the involvement forms of instruction and allocation of resources: Career Beginnings (Cave & Quint, 1990), the Cities in School federal demonstration (C.A. Murray et al., 1980), the Peer Tutoring and Mentoring Project (Turkel & Abramson, 1986), and Project RAISE (McPartland & Nettles, in press). The PUSH-EXCEL federal demonstration (S.R. Murray et al., 1982) was characterized by all four forms of involvement. Also included among the studies reviewed is an evaluation of a group of school-based clinics, which are not connected to each other in a formal sense (Kirby et al., 1989), but which feature several forms of resource allocation.

The instructional programs targeted elementary school students, whereas the other programs targeted students in elementary and secondary school. Although the studies are diverse in type and in the populations sampled, the sample can provide an overview of the kinds of effects that are possible for programs that have significant input from community entities.

Methodological Considerations

Community involvement for many years was viewed as good in and of itself. With the introduction of well-funded, multisite efforts and the concomitant inclusion of community involvement as a topic in discourse about school reform, program evaluation techniques were applied to assessments of community efforts. Although the literature still includes numerous examples of assessments the results of which are either ambiguous or so restricted as to be of use only for anecdotal purposes, the studies reviewed attempted to address methodological problems that are predictably associated with community involvement programs. Four of these issues—selection bias, fluctuations in sample composition, level of exposure to the treatment, and measures used—are discussed briefly below.

Selection bias. Selection bias is a particular threat in research to community-related interventions, because students typically can select themselves as program participants. Such selection can be tied to a number of factors that may render

TABLE 1
Study summary

	Cambridge-Somerville Youth Study (1937–1945)	Chicago Area Project	Pregnancy Prevention Program	Project Redirection demonstration (1980–1983)	Resource Mother Home Visit Program
Study citation	Powers & Witmer, 1951	Schlossman et al., 1984	Zabin et al., 1986	Polit & Kahn, 1985	Unger & Wandersman, 1985
Location	Cambridge, MA; Somerville, MA	S. Chicago, IL	Baltimore, MD	Boston, New York, Phoenix, Riverside, CA	Three rural counties in S. Carolina
Features	Dr. R. C. Cabot established foundation to provide paid counselors to "problem" boys. Caseload: 34–35 boys per counselor. Counselors did a wide variety of things: tutoring, arranging employment services, going on field trips, etc.	Two community-based organizations: The Mexican Community Committee (MCC) & S. Chicago Organized for People's Efforts (SCOPE) provided services for targeted youth & undertook community improvement & organization activities to prevent area delinquency.	Social worker & nurse, based in each of 2 schools, provided school-based services (e.g., counseling, homeroom presentations) & after-school services (educational & medical) in a clinic located in the community.	Implemented by community-based organizations according to design & guidance of Manpower Demonstration Research Corporation; provided social & employment services to teens, use of individualized participant plans, of community women, & of peer-group meetings	Paraprofessionals, called Resource Mothers, visited teen mothers once each month during the pregnancy and baby's 1st yr. Resource mothers used a structured curriculum to provide information and other services.
Sampling strategy	Random assignment to program & treatment groups of 325 matched pairs of boys	Program participants' addresses were matched to S. Chicago neighborhoods to identify specific areas served by the	Nonrandom; 2 program schools (1 junior high, 1 senior high), each paired with a control school	Nonrandom. Program sites paired with comparable cities; comparison teens recruited in same way as program partici-	Random assignment of teens expecting 1st child to Home Visit group (n = 70) & to comparison group (n = 17); compari-

		programs. Remaining S. Chicago neighborhoods served as comparisons: two program neighborhoods & 4 comparison neighborhoods.		pants (word of mouth & agency referrals), using same eligibility guidelines	son group members contacted once every 3 mos. by Resource Mothers and provided referral, but no services
Sample characteristics & attrition	Median age: 10.5 yrs; matched on physical health, IQ, delinquency prognosis, personality, home, & neighborhood environments. Of treatment group: 1 in 11 were Black; 163 exhibited delinquent behavior; 1/3 were extremely poor. Analysis was based on 148 pairs available after attrition in both samples (71 in E, 101 in C).	Percent of minorities in neighborhood ranged from 78–99; minority population aged 5–17 yrs. ranged from 25–32%; median rents in 1980 ranged from $122–$218	Program schools located in low-income community; control schools were racially mixed; only Black students used as comparisons; high baseline levels of sexual activity in program & nonprogram schools	789 in original sample; 675 remained after 2 years (93% of comparison group & 78% of experimental group). Sample was extremely disadvantaged: Approximately half not in school, 46.5% Black, 24% Mexican American, 17.1% Puerto Rican, & 9% White, nonHispanic. Comparison group was more likely to be married & enrolled in school.	Population was 89% Black, 11% White; 93% single; average age was 15.9; all were of low socioeconomic status.
Measures	Official court records; police contacts; counselor reports of social adjustment	Discrepancy between predicted & reported rates for 3 types of delinquency: runaway & ungovernable behavior, police contacts	Knowledge of contraceptive & pregnancy issues; contraceptive behavior; pregnancy rates. Data was obtained from self-administered	From interviews conducted at entry, at 12 mos., at 24 mos.: pregnancy status, contraceptive use, subsequent births, employment	Baby's birth weight; knowledge about babies; persistence in school; parenting behavior as measured by HOME

continued on p. ***IV-10***

TABLE 1 (Continued)

	Cambridge-Somerville Youth Study (1937–1945)	Chicago Area Project	Pregnancy Prevention Program	Project Redirection demonstration (1980–1983)	Resource Mother Home Visit Program
		(male vs. female), & arrests/contact (male vs. female)	questionnaires.		
Level of exposure to treatment	In 1 program (1940), counselors made an average of 27.3 visits/interviews per boy. Boys were in the program a minimum of 5 yrs.	N/A		42% of program teenagers were in program 12 mos. or more; 54% of comparison group had ever participated in a program for teen parents; most favorable sustained outcomes were found for participants in the program for 12 mos. or more; least favorable outcomes were for comparisons never in a program	11.9 Resource Mother contacts in visited groups; 5.7 contacts in comparison group

	Cities in Schools (federal demonstration)	Peer Tutoring & Mentoring Project	Project RAISE	PUSH for Excellence (federal demonstration)
Study citation	Murray et al., 1980	Turkel & Abramson, 1986	McPartland & Nettles, in press	Murray et al., 1982
Location	Atlanta, Indianapolis, New York	New York, NY	Baltimore, MD	Chattanooga, Denver
Features	Approximately 2,500 elementary & secondary students provided social, academic, & other services in groups	Ninth-grade, potential dropouts met with college student mentors over 2.5 mos. in sessions of 60–70	In progress; 7 groups of students provided mentoring or other support from adult volunteers; each	Schoolwide programs that featured inspirational speeches, incentives for achievement, enrichment

	of 40 students, called Families. Each family was staffed by 4 caseworkers, typically from agencies outside the school.	mins. Meetings were during/after school. Academic tutoring & personal school & social concerns were discussed.	group sponsored by a community business, university, fraternity, or church. Sponsors provided for each group a school-based adult advocate.	activities, attendance competitions, counseling, parent & community events
Sampling strategy	Nonrandom to program & comparison groups; randomization unsuccessfully attempted in 1 site; comparison groups consisted of nonprogram students in same school & grade as program participants	Nonrandom; comparison group consisted of 9th grade potential dropouts in same school as participants	Nonrandom; treatment groups consisted of students self-selected or selected by sponsors	Nonrandom. Students in sample were those whose parents gave permission to participate in the study. Comparisons were between levels of treatment.
Sample characteristics & attrition	487 in 1st yr. & 271 in 2nd yr.; 392 in comparison group in 1st yr.; 365 in 2nd yr. (All students were 9th graders in 1st yr. study.) Program samples were from lower income families than those of comparison group.	16 mentors; 18 comparison students	Treatment sample included all 7th-grade participants ($N = 314$); comparison group included 7th-grade students in 7 middle schools; students from both samples drawn from poorest neighborhoods in city	499 students in 7 schools comprised original sample; 2 schools dropped, due to extremely small sample sizes; remaining school samples experienced attrition rates from 11.7%–52%
Measures	Grades; attendance; Gates-McGinty Reading Test (Indianapolis) Iowa Test of Basic Skills (Atlanta), Nowick-Strickland Locus of Control Scale; Georgia Criterion Test (students' preparation for future)	Quality of School Life Scale (QSL); grade point average; attendance; standardized citywide reading test (not specified)	Grade point averages overall & in English & math; California Achievement Test scores (grade equivalents) in reading & math; absence & promotion rates	Individual: interview measures of personal efficacy; participation in extracurricular activities; academic self-concept & school-related behaviors; grades & attendance Schoolwide: attendance, suspensions
Level of exposure to treatment		Average of 6.5 meetings per mentor pair; range: 4–10 sessions	Not measured	Average number of activities per students ranged from 1.3–2.5 in program schools

continued on p. IV-12

TABLE 1 (Continued)

	School-based clinics	EXTRA	Parent-Child Tutoring Program	Career Beginnings
Study citation	Kirby, Waszak, & Ziegler, 1989	Sheley, 1984	Mehran & White, 1988	Cave & Quint, 1990
Location	Gary, IN; Muskegon, MI; Jackson, MI; West Dallas; Quincy, FL; San Francisco	Major southern city	Small western city (pop. 35,000)	Bronz, NY; Gary, IA; Indianapolis; IA; Jacksonville, FL; Rochester, NY; Santa Ana; Youngstown, OH
Features	Clinics provided medical services (physicals, lab tests, immunizations); counseling & health education; reproductive & family planning	Staff consisted of 5 paid public school teachers; sessions were M–F for 2.5 hrs.; students received help with homework & basic skills; open enrollment with some teacher referrals. 48 of 65 participants were from 1 school.	Parents trained in tutoring techniques using *Reading Made Easy;* asked to tutor students for 15 minutes, 3 times/wk. for 9 mos. (180 min./mo.); asked to attend monthly & bimonthly parent follow-up meetings & maintain logs	Counseling; summer jobs; workshops on college options; adult mentors from community
Sampling strategy	Nonrandom student survey; each school in 4 sites paired with a comparison school; in the remaining sites, clinics were not open at the outset of study, thus pre- and postclinic surveys were conducted; in 4 of 10 schools in study, stratified sample of classes was selected; in remaining schools, entire student body was sampling frame	Nonrandom; participating students from 1 school matched with controls according to age, gender, grade, level, teacher, course grades, & CTBS scores	Random assignment of pairs of students, matched according to CTBS scores, to program & comparison groups	Random assignment of 1,574 eligible, 11th-grade students to experimental & control groups

Sample characteristics & attrition	Samples ranged from 24%–90% of the school populations; sample sizes from 317–1,177. Analysis was restricted to Black students in all schools except San Francisco. Average age among Black students ranged from 15.7–16.8. Less than 16% in each sample was on welfare.	36 matched pairs in upper & lower grades; program participants had higher math scores in 57% of matches & higher reading scores in 52%	76 1st-grade students eligible for Chap. I compensatory education programs, identified in kindergarten; data available for 66–69 of children who remained after others were eliminated due to moving or absenteeism & scheduling problems	31.1% of sample received cash welfare; 68.9% Black, 14.5% Hispanic, 9.2% White, 7% Asian or Pacific Islander; 67.2% had parents that did not attend or graduate from college; 78% responded to 2 interviews (at high-school graduation & 1 yr. later)
Measures	Student Health survey; utilization from clinic records; school birth rate data; days absent due to illness	Reading & math subscales of CTBS, preprogram, after short term of enrollment & after long-term exposure to program	Harrison Criterion Referenced Test of Basic Skills (pre-, post-, & delayed posttests) Reading subscales of the CTBS (pre-, post-, & delayed posttests) Reading subscales of the Woodcock-Johnson Psycho-Educational Battery	College attendance & persistence; employment; productive activity; family formation; having children
Level of exposure to treatment	Approximately 60–80% of survey students in 4 sites used the clinics in a single year; across sites, half visited 3 times or less.	Students in sample averaged 4 days of attendance per wk.: 13 of 36 students in sample were enrolled for an average long term of 19 mos.; remainder enrolled for short term averaging 5.6 mos.	33% of parents tutored for average of 7 min./mo.; 19% tutored for average of 44 min./mo.	Significantly higher percentage of experimentals than controls received services during the evaluation period; ⅕ of the experimentals did not participate in the program

comparison groups of nonvolunteers ipso facto nonequivalent. For example, students who elect to participate may be motivated to take advantage of opportunities (or have parents who are unusually motivated to return forms giving their children permission to participate). Moreover, program operators are often reluctant on ethical grounds to assign randomly to treatment and control groups the students that seek services or are otherwise eligible for them.

In the studies reviewed, various approaches were used to control for possible initial differences between treatment and comparison groups. Random assignment to experimental and control groups from a pool of students eligible for the program was achieved in three of the studies. These included the two action research studies (the Parent-Child Tutoring Program and the Resource Mother Home Visit Program), where the investigators had authority over the design and implementation of both the program and the research, and the study of Career Beginnings, in which program administrators agreed to comply with the requirements for a rigorous research design.

To achieve control of extraneous and other unwanted variance in evaluations of programs whose recruitment and selection procedures were nonrandom and unrelated to the research designs, investigators used such approaches as comparison sites (Project Redirection), comparisons between levels of participation in the program (PUSH for Excellence), and multiple comparison groups, each designed to hold constant a given factor, or factors (McPartland & Nettles, in press). Many of the studies used statistical controls (such as baseline values of the dependent variables) as well.

Changes in sample composition. Most of the studies that collected data at more than one point in time, or at some point subsequent to the formation of treatment and comparison groups, were affected by the loss of participants from groups that constituted the original research sample. Two of the studies (Project Redirection and Career Beginnings) reported the results of tests of the representativeness of the responders to the original sample. There were significant differences in pertinent variables between respondents and nonrespondents in the Career Beginnings study, but the investigators concluded that the evaluation results were representative of a broad section of the original sample. The Redirection study used a two-stage statistical procedure to analyze the effects of attrition and found them to be negligible.

The foregoing examples addressed the issue of sample reduction due to respondent decisions. A related consideration is attrition due to program decisions about which students should remain in a program. It is not uncommon for program operators to reduce the numbers of participants to achieve a higher quality or a more manageable level of services. Such selection, which can violate the most elegant of research designs, occurred in the Project RAISE study, and the investigators therefore conducted separate analyses that compared the results before and after the student rosters were reduced.

Level of exposure to the treatment. Level of exposure to treatment is an issue for research design and analysis in that the kinds of services and activities that constitute a formal program are often available in other settings to members of potential and actual comparison groups. For example, in the PUSH for Excellence study, a major component was motivational speeches by a highly visible and charismatic public figure. Not only did the leader present speeches in program schools but he was also featured on a number of nationally televised talk and news shows.

In the Project Redirection and Career Beginnings studies, respondents in the comparison groups reported levels of the use of services comparable to those of program participants. The investigators addressed these unanticipated findings as issues in the interpretation of the results. Thus, in the Project Redirection study, the original question of the evaluation, which was whether any intervention was effective in assisting teenaged mothers, was modified to ask whether Project Redirection was more effective than other service models.

Measures. In their assessment of the potential of community-based, after-school literacy programs, Davidson and Koppenhaver, authors of the Project on Adolescent Literacy (1988), commented that measures of success for these programs differ from the standards used in school programs:

> Schools must attend to group objectives and standards, but after-school programs are free to focus intensively on individual goals. After-school programs deem themselves successful when they can engage a young person on a continuing basis, promote success in some area of learning, excite interest in some aspect of reading or writing, and help the individual to see that literacy does have a place in his or her future. (p. 132)

These authors recommended the use of qualitative measures of success that use as evidence such sources as program attendance records, student journals, and structured observations and interviews.

The studies in the present sample relied on measures that were used in evaluations of school and other institutional programs (e.g., school attendance, grades, standardized test scores, police contacts). Interviews were used to capture data on aspirations and plans, sense of efficacy, contraceptive use, and other self-reports, but structured observations were rare.

The different kinds of measures may be grouped into three categories. The first is *investments,* defined as students' commitments of their time, energy, and other resources in pursuit of legitimate opportunities that will yield a future return (Schwarz, 1980; Nettles, 1989). The focal investments in the studies reviewed were attending school, using contraception, participation in extracurricular activities, and working after school, or part-time. The second category includes measures of *attitudes,* such as sense of efficacy and attitudes toward school. The final category embraces measures of *attainment,* such as test scores, grades, promotion, and school completion.

Virtually all the studies addressed student engagement in the program through measures of level of exposure to the treatment. As Table 1 shows, the programs varied widely in the level of continuing participation, or use of services.

Effects on Individual Students

The programs sought positive changes in students enrolled as official participants and recipients of services, in entire schools, or in entire neighborhoods. The following section addresses the findings at the individual level of analysis.

Academic outcomes. Positive effects on reading skills were sought in five programs: each of the tutoring programs, Project RAISE, and Cities in Schools. Although significant improvement from pre- to posttest was found for program participants in the Cities in Schools evaluation, in the absence of a comparison group, factors other than the program may have influenced the results. The findings of the study of the Parent-Child Tutoring Program indicated that substantial gains on all measures were obtained for students whose parents participated at planned levels. For example, among pairs in which the experimental parents implemented the tutoring according to design, the average effect size was .96 across three tests (the Comprehensive Test of Basic Skills, the Woodcock Johnson Psycho-Educational Battery, and the Harrison Criterion Referenced Test) approximately nine months after the intervention. No significant differences were found between program participants and comparisons in the remaining four studies.

Gains in mathematics skills, which were the goals of two programs (Project RAISE and EXTRA), were obtained only in the EXTRA program. Students in the tutoring program showed greater improvement on the math subscale of the Comprehensive Test of Basic Skills in 64% of pairs matched according to age, grade level, course grades, test scores and teacher. However, matched-pair differences in both reading and mathematics were related to gender and tenure in the program. Among long-term pairs, girls in the program showed greater improvement over their control twins than boys did with respect to their matches. Differences in mean score changes were 26.67 and 83.29 for boys and girls respectively in math and 14.25 and 78.00 for boys and girls respectively in reading.

Improvements in grades were examined in the studies of Peer Tutoring and Mentoring, Project RAISE, Cities in Schools, and the PUSH-EXCEL Project. In the latter, level of participation in the program was positively associated with grade point average after the effects of preprogram grade point average were taken into account. Students in Project RAISE received better English (but not math grades) than other students in the same schools. The size of the positive program effect was .14. However, the RAISE students' English grades remained below the average for the school district. In addition, there were no significant differences in grade point averages between program participants and comparisons in the evaluation of the Peer Tutoring and Mentoring project.

Attendance. Improved school attendance was sought in four programs (Peer Mentoring and Tutoring, Cities in Schools, the PUSH-EXCEL Project, and Project RAISE). In the studies of Project RAISE and Cities in Schools, significant differences in the anticipated direction were found between comparison students and program students in the same school and grade. For example, the reduction in absences due to participation in RAISE was nearly 3%, which translates into about one week of extra days of attendance in a 180-day school year. Among students in one of the cities in the Cities-in-Schools program, there was nearly a 7% decrease in absences in the 8th to 9th grades compared to no decrease among comparison students. No significant effects were found in studies of the other two programs.

Persistence in school. Five of the programs (Cities in Schools, Project Redirection, Project RAISE, Career Beginnings, and the Resource Mother Home Visit Program) sought to induce students to remain in school or to make satisfactory progress toward graduation from high school or from a postsecondary program. The evaluations of three of the programs found that the programs were effective in this regard. The Project Redirection study found that, 12 months after the program began, a significantly higher proportion of the participant group than of the comparison group was in school or had graduated (56% compared to 49%). At eight months postpartum, a higher proportion of mothers in the visited group of the Resource Mother program remained in school than their counterparts in the comparison group, and a significantly higher percentage of participants (47.9%) than of controls (43.4%) in the study of the Career Beginnings program was in college after one year. The results for these three studies are particularly credible because the research designs in the evaluations were stronger than in the others reviewed here.

Other short-term effects. The studies also produced evidence that community-initiated, or community-operated, programs can produce desirable effects on student attitudes toward the self and school and on risk-taking behavior. For example, a heightened sense of personal efficacy was associated with participation in the Cities in Schools and the PUSH-EXCEL Project evaluations. Students in the Peer Tutoring and Mentoring Project had more positive attitudes toward school (as measured by pre- and postprogram schools on the Quality of School Life Scale).

In the evaluation of school-based clinics, in three sites in which clinic users were compared to clinic nonusers and where clinics made contraceptives available to students by dispensing them or by providing vouchers, students who elected to use the clinics for contraception were more likely to have used contraceptives at last intercourse than students who did not use the clinic for this service. Similarly, in the study of Project Redirection, one year after the program began, a greater proportion of participants than of comparison subjects reported contraceptive use at last intercourse. Moreover, during the first year of the study, a significantly lower percentage of project participants than of comparison subjects had a repeat pregnancy. There were no differences in employment status between the groups at 12 months, although at the 12-month interview the proportion of participants ever employed was greater than that of comparison subjects ever employed.

Long-term effects. Of the 13 studies reviewed, three of them (the evaluations of the Cambridge-Somerville Youth Project, of Career Beginnings, and of Project Redirection) measured effects subsequent to respondents' participation in program activities. The effects on participants' persistence in college achieved by the Career Beginnings program were mentioned above. In the Projection Redirection study, measures were taken at 12 months, when the average participant was ending her involvement in the program, and at 24 months. Overall, differences found at 12 months between comparison participants and program participants in subsequent pregnancy, employment status, and school enrollment, or school completion had vanished by 24 months. Exceptions to this were found in analyses of subgroups: Desired effects were sustained for program participants who were extremely disadvantaged relative to other participants.

The longitudinal study of the Cambridge-Somerville Youth Project suggests that intervention models may provide short-term, but not continuing, amelioration of the problems associated with disadvantage. Because this insight is essential for understanding the effects of interventions, the program and the original research design are briefly described.

Dr. Richard Clark Cabot initiated the Cambridge-Somerville Youth Project in Massachusetts in 1935 in an effort to curb delinquency. The police, churches, schools, and social service agencies recommended boys aged 5 through 13 fo participation in the project. Boys considered to be "average" and "difficult" were identified and paired according to delinquency prediction scores and personal and home background. One member of each pair was assigned to the control group and the other to the treatment that began in 1939. For approximately 5 years, the project arranged for academic and medical services as needed, linked the boys to youth and other community organizations, and sent one fourth to summer camp. Twice a month, counselors visited the boys' families. The initial evaluation of the Cambridge-Somerville Youth Study found no significant differences in official delinquency or in social adjustment between treatment and control groups (Powers & Witmer, 1951). However, among pairs whose treatment twin received services that removed barriers to adequate socialization, the treatment group had a higher level of social adjustment.

In 1975, McCord (1978) traced the whereabouts of 506 men who had participated in the treatment and control groups. Of the 480 who were located, 48 had died. Questionnaires were sent to the remainder (208 in the treatment and 202 in the control) and data from the Massachusetts archives on the 340 men still living in the state were examined.

The men in the treatment group rated the program very positively, but the analyses indicated that the program had no subsequent impact on delinquency. Moreover, the participants in the treatment group, compared to those in the control group, tended to (a) show more signs of mental illness, (b) have had at least one stress-related disease, (c) be in low prestige occupations, (d) show signs of alcoholism, (e) report more often that their work was less satisfying, and (f) have committed a second crime. There were no statistically significant differences between the treatment group and the control group in 50 other comparisons.

Subsequent analyses (McCord, 1981) indicated that participants who had particularly long, early, or frequent contact with the counselors showed the strongest adverse impact. After testing several interpretations of the effects (e.g., that the program increased dependency), McCord concluded that the program

> seems to have raised the expectations of its clients without also providing the means for increasing satisfactions. The resulting disillusionment seems to have contributed to the probability of having an undesirable outcome. (p. 405)

Effects on School and Area Populations

In 4 of the 13 programs, the entire school, or entire neighborhood, was the target of intervention, rather than students who were recognized as official participants. The program included the PUSH-EXCEL Project, the Chicago Area Project, the Pregnancy Prevention Program, and six school-based clinics.

School attendance. Changes in school attendance were examined in the studies of the PUSH-EXCEL Project and the school-based clinics. In the five schools in the PUSH-EXCEL study, absences decreased slightly (1–2 percentage points) in all schools concurrently with the implementation of the program. In three of the sites in the evaluation of school-based clinics, there were no significant differences found in days absent due to illness in clinic schools versus comparison schools that did not have clinics. However, in two schools that opened clinics after the evaluation began, students missed fewer days after the clinics opened than they had before the clinics opened.

Pregnancy prevention and reduction of high-risk behaviors. The evaluation of the Pregnancy Prevention Program assessed changes over time in students' sexual behavior, contraception, pregnancy rates, and knowledge and attitudes about reproductive issues. The investigators used data from school archives and from surveys conducted in the two project schools and in the two control schools. Young women who attended the school concurrently with the three years of program implementation initiated sexual intercourse a median length of seven months later than those who attended the school prior to program implementation. In program schools, as compared to control schools, there were reductions in rates of unprotected intercourse and pregnancy.

The evaluation of six school-based clinics found that the clinics had no impact on schoolwide pregnancy and birth rates. Nevertheless, at two of the sites, significantly greater proportions of students in clinic versus nonclinic schools (75% compared to 61%) reported using contraceptives at last intercourse. In addition, significantly greater proportions of students in pre-clinic schools (66% before the clinic opened compared to 75% 2 years after the clinic opened) reported using contraceptives at last intercourse. Alcohol consumption and cigarette smoking were measured in four sites. Students in three clinic sites reported lower alcohol consumption, and students in one of the four clinic sites reported lower cigarette smoking than students in nonclinic schools.

The evaluation of the Chicago Area Project examined areawide reductions in three types of delinquency: runaway and ungovernable behavior, police contacts, and arrests. The analysis compared the discrepancy between predicted and reported rates in six neighborhoods: two program neighborhoods (identified by matching individual participants' addresses to specific areas in South Chicago) and four comparison neighborhoods (the remaining neighborhoods in South Chicago). In one program neighborhood, all measures of delinquency were lower than expected. In the second program neighborhood, delinquency was lower than expected on all measures except male arrests and police contacts. Two comparison neighborhoods showed higher rates than expected, and two (one of which was served by a community center not in the program) showed lower rates.

Summary

Table 2 summarizes the direction of short-term effects found in the studies reviewed according to form of involvement. It is clear that the programs can have positive effects on school-related behaviors and achievement as well as on attitudes and risk-taking behavior. Within types of effects, the consistency of positive outcomes for attendance, pregnancy status and contraceptive behavior, and persistence in school suggests that community programs may be potentially useful interventions. Overall, the effects range from small to substantial. This is not surprising given the variations in level of exposure to treatment and quality of research design.

Also of note is the pattern of outcomes by involvement types. Programs that fall either in the allocation or the instruction categories tend to show an overall pattern of positive effects. Programs that combine allocation and instruction show a mixed pattern. There is only one program that combined the four forms of involvement, and thus there is no basis for comment on patterns.

Directions for Future Research

The literature on the effects of community involvement on disadvantaged students is predominantly a literature of program evaluation. The studies reviewed above should be considered in light of their occurrence in the history of evaluation research. The critiques (Farrar & House, 1983; Stake, 1983) of the earliest studies of grassroots involvement, the Cities in Schools and the PUSH-EXCEL Project evaluations, suggested the need for evaluators to respect the essential and unique character of a given intervention, to employ compatible methods in assessing effectiveness, and to

TABLE 2

Summary of outcomes by form of community involvement

	Academic					
	Test scores	G.P.A.	Attendance	Persistence in school	Pregnancy status & contraceptive use	Other
Allocation (A)			School-based clinics Schoolwide 2 of 5 sites + Remainder ○	Project Redirection + Resource Mother Home Visit +	Increased use of contraception Pregnancy Prevention Program Schoolwide + Project Redirection + School-based clinics Individual clinic users + Schoolwide 2 of 5 sites + Remainder ○ Decreased pregnancy Pregnancy Prevention Program Schoolwide + Project Redirection + School-based cinics Schoolwide ○	Decreaed delinquency Cambridge-Somerville ○ Chicago Area Project + Decreased cigarette smoking School-based clinics 1 of 4 sites + Remainder ○ Decreased alcohol consumption School-based clinics 3 of 4 sites + Remainder ○
Instruction (I)	EXTRA Reading ○ Math + Parent-Child Reading +					
Allocation & instruction (AI)	Cities in Schools Reading + Peer Tutoring & Mentoring Reading ○ RAISE Reading ○ Math ○	Cities in Schools ○ Peer Tutoring & Mentoring ○ RAISE English + Math ○ Overall ○	Cities in Schools + Peer Tutoring & Mentoring ○ RAISE +	Career Beginnings + Project RAISE ○		Heightened sense of efficacy Cities in Schools + Improved attitudes toward school Peer Tutoring & Mentoring +
All (C, M, A, I)		PUSH for Excellence +	PUSH for Excellence Individual ○ Schoolwide +			Heightened sense of efficacy PUSH for Excellence +

○ = n.s.d.

+ = program participants or schools (schoolwide) significantly higher (in desired direction) than controls maintain clear boundaries between evaluation and formulation of program policy. Recent studies, such as the evaluation of Project Redirection and the evaluation of school-based clinics, reflect advances that have been made in design analysis, in data analysis, and in the increased cooperation among program implementers, funders, and evaluators in the planning and implementation of research. The establishment of evaluation networks (e.g., the National Network of Teen Pregnancy Programs Doing Impact Evaluations), information on evaluation targeted at program operators (see Philliber, 1989, for an illustrative handbook), and the application of approaches that contribute to program development (cf., Gottfredson, 1984a) bode well for continuing progress in the study of the effects of community involvement.

However, a major weakness in the existing evaluation research is the lack of attention to the effectiveness of components of programs. Research is needed to help program developers confront two difficult challenges: identifying effective practices from among the scores of programs that now exist and fostering student participation in program activities. Studies to increase participation levels, for example, might focus on structuring incentives that meet the following criteria: Incentives should be appropriate in terms of the students' developmental level, abilities, and interests; incentives should be inexpensive; and incentives should not undermine program goals or community and family norms, values, or resources.

Two major gaps are apparent in the general field of community involvement research. One is the dearth of studies on the quality and effects of naturally occurring and institutionalized occurrences of the four forms of involvement. Also needed are studies that answer questions about the relationship between intervention and informal involvement. For example: Does planned involvement facilitate or impede naturally occurring community processes? What is the nature of informal instruction in local businesses, churches, settlement houses, and youth organizations? Can planned support improve informal practices? What factors stimulate, or impede, involvement?

The second major problem with the existing research on community involvement is its conceptual isolation from research on how communities and other ecologies affect disadvantaged students. The programmatic efforts and studies included here were guided by highly specific practical or theoretical rationales, rather than by a general conceptual framework. These specific rationales are reflected in the narrow range of questions answered in the research, the focus on indicators of school success and on adjustment to the exclusion of other measures, and the absence of information that would be helpful in designing effective treatments and strengthening informal practices. This article proposes a conceptual framework to assist both practitioners and researchers in integrating existing studies in varied disciplines, in sketching an outline for new lines of inquiry, and in identifying new directions for the design of interventions.

This conceptual framework integrates three separate lines of research. The first assesses *community competence,* which refers to the capacity of a community and the agents within it to solve problems and to meet the demands of daily life (Barbarin, 1981; Iscoe, 1974). Communities that function well are in some respects the counterparts of effective schools. This research literature suggests that competent communities are characterized by such features as responsiveness to the diverse needs of members, maximized use of resources, cohesiveness, and a collective sense of well-being, of physical security, and of opportunities for individuals to achieve status and receive recognition for accomplishments. These and other characteristics can be defined as one of three components—namely:

1. community structure, which embraces physical features, social area characteristics, and other aspects of the community's resources;
2. community culture, or climate, which is defined by values, standards, and rules; and
3. community processes such as problem solving and allocation of resources (Hurley, Barbarin, & Mitchell, 1981).

This framework encompasses the two most common meanings of *community*. One refers to *community* as a locality, such as the neighborhood, the city, the block, or the catchment area of a school. The other meaning views *community* as the social interactions that occur in formal and informal settings within, and across, locales. (See Heller, 1990, and Newmann & Oliver, 1969, for discussions of these meanings and their implications.) Understanding these dimensions of community will contribute to research in three ways: It will help address the neglect of processes that occur naturally in communities, aid in the specification of connections between community characteristics and community involvement, and give researchers insight into how local variation, which has not been addressed systematically in the literature, shapes the direction and intensity of involvement.

The second line of research evaluates the influence of *educational environments* on student development—how the structure, social climate, and processes of classrooms, schools, dormitories, families, and other institutional settings affect student performance, aspirations, attitudes toward school, delinquency, and other behavioral and cognitive outcomes (Astin, 1968; G. D. Gottfredson, 1984b; Moos, 1979). In Moos' (1979) model, students' cognitive, motivational, and coping mechanisms link the environmental system (i.e., structure and climate dimensions) and the student's personal system (i.e., sociodemographic characteristics, personality, and skills) to changes in student values, interests, aspirations, and achievement levels. Thus, studies of community involvement should account for student characteristics (beyond economic status or academic achievement, two common factors that are already used to distinguish subgroups) that may mediate the effects of involvement.

The third line of research includes studies of involvement, such as the evaluations reviewed above, as well as case studies and other forms of research. These studies have begun to specify the important student variables that can be influenced by the actions of community entities.

Figure 1 shows the proposed framework. It defines the community as an environment characterized by three measurable features: structure, culture or climate, and the involvement process. Structure and climate are aspects added to the typology of community involvement developed previously and used throughout this article. *Community structure* refers to the nature and organization of the social units and physical features within the community's boundaries. Four dimensions of community structure dominate in the literature as important targets of involvement or as factors affecting student outcomes: the educational resource base, history, social area characteristics, and the physical setting. Several studies indicate that structural characteristics have a direct effect on student attainment (for reviews, see Mayer & Jencks, 1989; Scott-Jones, 1989). The line connecting structure to involvement is suggested by case studies of partnerships and citizen participation, which indicate

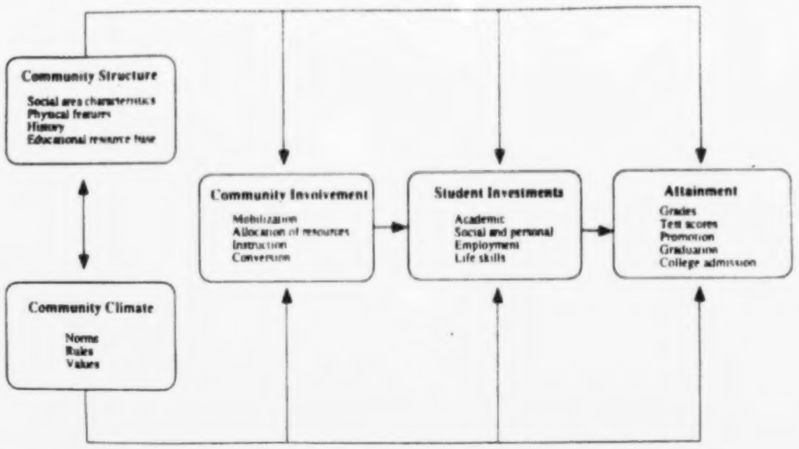
Community Structure
Social area characteristics
Physical features
History
Educational resource base
Community Climate
Norms
Rules
Values
Community Involvement
Mobilization
Allocation of resources
Instruction
Conversion
Student Investments
Academic
Social and personal
Employment
Life skills
Attainment
Grades
Test scores
Promotion
Graduation
College admission

FIGURE 1. *Framework for examining community involvement and student progress*

that community history and the level of resources shape the form and direction of involvement.

Community climate consists of the values, norms, and rules that serve to maintain community order and control, to promote extensive social interaction among community members, and to facilitate individual community members' growth and progress. This aspect of community has been explored largely through ethnographic studies (e.g., Lightfoot, 1978; Anderson, 1976). These studies may be useful starting points for operationalizing cultural elements that influence involvement and student development. With regard to the latter, Ogbu's (1985) cultural-ecological model of inner city childrearing and development specifies the competencies that African Americans in the inner city expect children to acquire and the cultural factors that shape the type and content of such competencies. Thus, this model suggests that climate may have direct effects on student outcomes. Case studies of local organizing in the PUSH-EXCEL evaluation (Kumi, Thompkins, Allen, Murray, 1979) suggest that climate is an important influence on the level of mobilization for school improvement and, hence, on the line connecting climate to involvement.

Community involvement is conceptualized as the typology of the four involvement processes used in this review. These processes, singly and in concert, comprise the formal and informal actions that individuals and groups undertake either directly to foster student development or indirectly to improve or to reform institutions that serve youth. As this review suggests, the involvement of community actors can stimulate student investments such as attending school, using contraceptives, and avoiding high-risk behaviors such as alcohol consumption and delinquency. Attitudinal shifts and heightened achievement are also outcomes of some forms of involvement. However, different types of community involvement may not be equally effective in producing results.

Conceptualizing community in this way focuses attention on the aspects of community that affect the intellectual and psychosocial development of children and youths. By distilling what is already known about community environments and their effects on students, by implementing ambitious action research designs in program evaluations, and by exploring connections between the various aspects of community, investigators can contribute to practical and empirical knowledge about ways to remove impediments to the progress of disadvantaged students and can create environments that nurture these students.

References

Alinsky, S. (1971). *Rules for radicals.* New York: Random House.

Anderson, E. (1976). *A place on the corner.* Chicago: University of Chicago Press.

Asche, J. (1989). *Handbook for principals and teachers: A collaborative approach for the effective involvement of community and business volunteers at the school site.* Alexandria, VA: National Association of Partners in Education.

Ascher, C. (1988). *Urban school-community alliances.* New York: Columbia University, Teachers College. (ERIC Clearinghouse on Urban Education)

ASPIRA Association. (1990). *Latinos and the dropout crisis: The community solution.* Washington, DC: Author.

Astin, A. W. (1968). *The college environment.* Washington, DC: American Council on Education.

Bain, J. G., & Herman, J. L. (1989). *Improving opportunities for underachieving minority students: A planning guide for community action.* Los Angeles: UCLA, Center for Research on Evaluation, Standards, and Student Testing and UCLA, Center for the Study of Evaluation.

Barbarin, O. A. (1981). Community competence: An individual systems model of institutional racism. In O. A. Barbarin, P. R. Good, O. M. Pharr, & J. A. Siskind (Eds.), *Institutional racism and community competence* (pp. 6–19). Washington, DC: U.S. Government Printing Office.

Carnegie Corporation of New York. (1987–88). Black churches: Can they strengthen the black family? *Carnegie Quarterly, 33,* 1–7.

Carnegie Council on Adolescent Development. (1989). *Turning points: Preparing American youth for the 21st century, the report of the Task Force on Education of Young Adolescents.* New York: Carnegie Corp.

Cave, G., & Quint, J. (1990). *Career beginnings impact evaluation: Findings from a program for disadvantaged high school students.* New York: Manpower Research Development Corp.

Children's Defense Fund. (1986). *Preventing adolescent pregnancy: What schools can do.* Washington, DC: Author.

Constable, R., & Walberg, H. (1988). School social work: Facilitating home, school, and community partnerships. *Urban Education, 22,* 429–443.

Davidson, J., & Koppenhaver, D. (1988). *Adolescent literacy: What works and why.* New York: Garland.

Epstein, J. L. (1987). What principals should know about parent involvement. *Principal, 66* (3), 6–9.

Epstein, J. L., & Scott-Jones, D. (1988, November). *School-family-community connections for accelerating student progress.* Paper presented at the Centennial Conference of Stanford University, CA.

Farrar, E., & House, E. R. (1983). The evaluation of Push/Excel: A case study. *New Directions for Program Evaluation, 17,* 31–57.

Flaxman, E., Ascher, C., & Harrington, C. (1988). *Mentoring programs and practices: An analysis of the literature.* New York: Teachers College, Institute for Urban and Minority Education.

Freedman, M. (1988). *Partners in growth: Elder mentors and at-risk youth.* Philadelphia: Public Private Ventures.

Nettles

George, Y. S., Richardson, V., Lakes-Matyas, M., & Blake, F. (1989). *Saving minds: Black churches and education.* Washington, DC: American Association for the Advancement of Science.

Gottfredson, G. D. (1984a). A theory-ridden approach to program evaluation: A method for stimulating researcher-implementer collaboration. *American Psychologist, 39,* 1101–1112.

Gottfredson, G. D. (1984b). *The effective school battery.* Odessa, FL: Psychological Assessment Resources.

Gundlach, R., Farr, M., & Cook-Gumperz, J. (1989). *Writing and reading in the community.* Berkeley: University of California, Center for the Study of Writing.

Halpern, R. (1990). Community-based early intervention. In S. G. Meisels & J. P. Shonkoff (Eds.), *Handbook of early childhood intervention* (pp. 469–498). Cambridge, MA: Cambridge University Press.

Hargroves, J. S. (1986). The Boston Compact: A community response to school dropouts. *Urban Review, 18,* 207–217.

Heath, S. B. (1989). Oral and literate traditions among black Americans living in poverty. *American Psychologist, 44,* 367–373.

Heller, K. (1990). Social and community intervention. *Annual Review of Psychology, 41,* 141–168.

Henderson, A. (Ed). (1987). *The evidence continues to grow: Parent involvement improves student achievement.* Columbia, MD: National Committee for Citizens in Education.

Hurley, D. J., Barbarin, O. A., & Mitchell, R. E. (1981). An empirical study of racism in community functioning. In O. A. Barbarin, P. R. Good, O. M. Pharr, & J. A. Siskind (Eds.), *Institutional racism and community competence* (pp. 134–145). Washington, DC: U.S. Government Printing Office.

"I Have a Dream" Foundation. (1989). *The "I Have a Dream" program: Description and organization structure.* New York: Author.

Iscoe, I. (1974). Community psychology and the competent community. *American Psychologist, 29,* 607–613.

Kirby, D., Waszak, C. S., & Ziegler, J. (1989). *An assessment of six school-based clinics: Services, impact, and potential.* Washington, DC: Center for Population Options.

Kumi, L. M., Thompkins, N., Allen, T. W., & Murray (Nettles), S. R. (1979). *The national evaluation of the PUSH for Excellence project: Program descriptions.* Washington, DC: American Institutes for Research.

Levine, M., & Trachtman, R. (Eds.) (1988). *American business and the public school: Case studies of corporate involvement in public education.* New York: Teachers College Press.

Lightfoot, S. L. (1978). *Worlds apart.* New York: Basic Books.

Mann, D. (1987a). Business involvement and public school improvement, Part I. *Phi Delta Kappan, 69,* 123–128.

Mann, D. (1987b). Business involvement and public school improvement, Part II. *Phi Delta Kappan, 69,* 228–232.

Mayer, D. E., & Jencks, C. (1989). Growing up in poor neighborhoods: How much does it matter? *Science, 243,* 1441–1445.

McAdoo, H. P. (1980). Black mothers and the extended family support network. In L. F. Rodgers-Rose (Ed.), *The Black woman* (pp. 125–144). Beverly Hills: Sage.

McCord, J. (1978). A thirty-year follow-up of treatment effects. *American Psychologist, 33,* 284–289.

McCord, J. (1981). Consideration of some effects of a counseling program. In S. E. Martin, L. B. Sechrest, & R. Redner (Eds.), *New directions in the rehabilitation of criminal offenders* (pp. 394–405). Washington, DC: National Academy.

McPartland, J., & Nettles, S. M. (in press). A two-year evaluation of Project RAISE. *American Journal of Education.*

Mehran, M., & White, K. R. (1988). Parent tutoring as a supplement to compensatory education for first-grade children. *Remedial and Special Education, 9*(3), 35–41.

Merenda, D. W. (1986). *A practical guide to creating and managing school/community partnerships.* Alexandria, VA: National Association of Partners in Education.

Moos, R. H. (1979). Evaluating educational environments. San Francisco: Jossey-Bass.

Murray, C. A., Bourgue, B. B., Harnar, R. S., Hersey, J. C., Murray (Nettles), S. R., Overbey, D., & Stotsky, E. S. (1980). *The national evaluation of the Cities in Schools program* (Report No. 3: Program impact, 1978–79). Washington, DC: American Institutes for Research.

Murray (Nettles), S. R., Murray, C. A., Gragg, F. E., Kumi, L. M., & Parham, P. A. (1982). *The national evaluation of the PUSH for Excellence project: Final report.* Washington: American Institutes for Research. (ERIC Document Reproduction Service No. ED 240 225)

Natriello, G., McDill, E. L., & Pallas, A. M. (1990). *Schooling disadvantaged children: Racing against catastrophe.* New York: Teachers College Press.

Nettles, S. M. (1989). The role of community involvement in fostering investment behavior in low-income Black adolescents: A theoretical perspective. *Journal of Adolescent Research, 4,* 190–201.

Oakes, J. (1987). *Improving inner-city schools: Current directions in urban district reform.* Santa Monica: Rand.

Ogbu, J. U. (1985). A cultural ecology of competence among inner-city Blacks. In M. E. Spencer, G. K. Brookins, & W. R. Allen (Eds.), *Beginnings: The social and affective development of black children* (pp. 45–66). Hillsdale, NJ: Erlbaum.

One to One. (1990). *One to one: Mobilizing for the next generation.* Washington, DC: Author.

Orum, L. S. (1988). *Making education work for Hispanic Americans: Some promising community-based practices.* Los Angeles: National Council of La Raza.

Otterbourg, S. D. (1986). *School partnership handbook: How to set up and administer programs with business, government and your community.* Englewood Cliffs, NJ: Prentice Hall.

Philliber, S. (1989). *Evaluating your adolescent pregnancy program: How to get started* (Adolescent Pregnancy Prevention Clearinghouse Report). Washington, DC: Children's Defense Fund.

Pine, G., & Keane, W. (1989, March). *Collaboration for educational excellence and equity: Lessons learned.* Paper presented at the annual meeting of the American Association of Colleges for Teacher Education, Anaheim, CA. (ERIC Documentation Reproduction Service No. ED 304 437)

Polit, D. F., & Kahn, J. R. (1985). Project Redirection: Evaluation of a comprehensive program for disadvantaged teenage mothers. *Family Planning Perspectives, 17,* 150–155.

Powers, E., & Witmer, H. (1951). *An experiment in the prevention of delinquency: The Cambridge-Somerville youth study.* New York: Columbia University Press.

Rothman, J., Erlich, J. L., & Teresa, J. G. (1976). *Promoting innovation and change in organizations and communities: A planning manual.* New York: Wiley.

Rothman, J., Erlich, J. L., & Teresa, J. G. (1981). *Changing organizations and community programs.* Beverly Hills: Sage.

Schlossman, S., Zellman, G., & Shavelson, R. (1984). *Delinquency prevention in South Chicago: A fifty-year assessment of the Chicago area project.* Santa Monica: Rand. (ERIC Document Reproduction Service No. ED 266 195)

Schwarz, P. A. (1980). Program devaluation: Can the experiment reform? In E. H. Loveland (Ed.), *New directions for program evaluation* (pp. 65–91). San Francisco: Jossey-Bass.

Scott-Jones, D. (1984). Family influences on cognitive development and school achievement. *Review of Education Research, 11,* 259–304.

Scott-Jones, D. (October, 1989). *Families, communities, and schools as contexts for literacy and learning.* Paper presented at the conference on Literacy Among Black Youth: Issues in Learning, Teaching, and Schooling, University of Pennsylvania.

Sheley, J. F. (1984). Evaluation of the centralized, structured, after-school tutorial. *Journal of Education, 77,* 213–218.

Slaughter, D. T., & Epps, E. G. (1987). The home environment and academic achievement of Black American children and youth: An overview. *Journal of Negro Education, 56,* 3–20.

Stack, C. B. (1974). *All our kin.* New York: Harper & Row.

Stake, R. E. (1983). Stakeholder influence in the evaluation of Cities in Schools. *New Directions for Program Evaluation, 17,* 15–30.

Tangri, S., & Moles, O. (1987). Parents and the community. In V. Richardson-Koehler (Ed.), *Educator's handbook: A research perspective* (pp. 519–550). New York: Longman.

Turkel, S. B., & Abramson, T. (1986). Peer tutoring and mentoring as a dropout prevention strategy. *The Clearing House, 60,* 68–71.

Unger, D. G., & Wandersman, L. P. (1985). Social support and adolescent mothers: Action research contributions to theory and applications. *Journal of Social Issues, 41,* 29–45.

U.S. Department of Education. (1990). *College sponsored tutoring and mentoring programs for disadvantaged elementary and secondary students.* (Higher Education Survey Report No. 12). Washington, DC: Author.

Walker, G. (1984). Hispanic youth employment: Some lessons and models for business involvement. In R. Valdivieso & M. Galindo (Eds.), *Make something happen: Hispanics and urban high school reform: Vol. II* (pp. 79–83). Washington, DC: Hispanic Policy Development Project. (ERIC Document Reproduction Service No. ED 253 599)

Weisbaum, K. S. (1990). *Families in family math research project: Final report.* Berkeley: University of California, Lawrence Hall of Science.

Williams, M. R. (1989). *Neighborhood organizing for urban school reform.* Wolfeboro, NH: Teachers College Press.

W. T. Grant Foundation Commission on Work, Family and Citizenship. (1988). *The forgotten half: Pathways to success for America's youth and young families.* Washington, DC: Author.

Zabin, L. S., Hirsch, M. B., Smith, E. A., Streett, R., & Hardy, J. B. (1986). Evaluation of a pregnancy prevention program for urban teenagers. *Family Planning Perspectives, 18,* 119–126.

Author

SAUNDRA MURRAY NETTLES is Principal Research Scientist, Center for Research on Effective Schooling for Disadvantaged Students, Johns Hopkins University, 3505 N. Charles St., Baltimore, MD 21218. She specializes in evaluation of educational and social programs.

Using Community Adults as Advocates or Mentors for At-Risk Middle School Students: A Two-Year Evaluation of Project RAISE

JAMES M. MCPARTLAND and SAUNDRA MURRAY NETTLES
Johns Hopkins University

The effects on selected student outcomes are evaluated after two years of operation of Project RAISE, a multifaceted approach featuring outside adults as school-based advocates and one-on-one mentors for at-risk students at seven middle schools. Positive effects are found on improving student attendance and report card grades in English, but not on promotion rates or standardized test scores. The effects, though sizable, were not sufficient to neutralize the academic risks with which students entered the program. The positive results were primarily due to three of the seven sites. Some evidence supported interpretations that, although strong one-on-one mentoring is not an essential component of an effective program that uses outside adults to assist at-risk middle school students, the RAISE model is much more likely to show positive effects when one-on-one mentoring has been strongly implemented. Success may also depend on the size and composition of the student group to be served. Issues are raised about roles and responsibilities of adult advocates or mentors.

Schools throughout the nation are engaged in programs that use adults from the community to help at-risk youth make steady progress through the middle and secondary grades and complete high school. Two general approaches—mentoring and advocacy—are widely viewed as promising mechanisms to provide sustained, goal-directed support to students.

Mentoring is commonly defined as a one-to-one relationship between a caring adult and a student who needs support to achieve academic, career, social, or personal goals. Mentor-student relationships can develop naturally or within structured interventions through activities

© 1991 by The University of Chicago. All rights reserved.
0195-6744/91/9904-0009$01.00

designed to arrange, sustain, and monitor matches. Advocacy, as the term is currently applied, refers to a supportive relationship wherein a resourceful adult (who may be called an advocate, program coordinator, youth worker, or counselor) works with the same group of students over a specified period of time and provides intensive instrumental, material, and emotional support that can include assessing students' needs for academic and social services, intervening on the students' behalf in schools and other institutions, monitoring students' participation in programs, and identifying and brokering formal services.

Both of these approaches are extremely popular, not only for at-risk youth but for other populations as well. For example, a directory compiled by the New York State Mentoring Committee (1989) lists over 211 mentoring programs in New York State alone. A recent survey of college and university tutoring and mentoring programs for disadvantaged youth (U.S. Department of Education 1990) reported that, of an estimated 1,701 such programs, 63 percent provided mentoring and 17 percent had mentoring as the primary focus.

National mobilizations are under way to promote further the use of these approaches. For example, there is the growing number of activities under the "I Have a Dream" Foundation. As of October 1989, the program was under way in over 30 cities in 23 states (Berger 1989; J. M. Sesnick, personal communication, December 6, 1989). Another major effort is One to One, which has the goal of matching with a "caring partner" every young person who can benefit from such a relationship. This is to be accomplished through the formation of local leadership councils, pilot neighborhood projects, and the National Mentoring Partnership (One to One 1990). Big Brothers/Big Sisters of America has a long history of providing young people with adult volunteers in one-to-one relationships. The organization has nearly

---

JAMES M. MCPARTLAND is codirector of the Center for Social Organization of Schools and professor of sociology at Johns Hopkins University. His areas of interest include effects of school organization factors on student development and various issues of equality of educational opportunities. SAUNDRA MURRAY NETTLES is a principal research scientist in the Center for Research on Effective Schooling for Disadvantaged Students (CDS) and the Center on Families, Communities, Schools and Children's Learning at Johns Hopkins University. She is also currently codirector of the Consortium for Research on Black Adolescence. She has conducted several evaluations of programs that connect communities and schools.

August 1991

500 affiliates throughout the nation, and the children it serves include a large proportion from low-income or single-parent families (Smink 1990).

The widespread use of terms such as "mentoring" and "advocacy" and the prominence in the media of testimonials about how kids have been "turned around" through contact with a mentor or advocate give the impression that these are well-defined approaches that effectively increase students' motivation and achievement in school, remove barriers to student progress in school and the wider community, and help students refrain from self-destructive and illegal actions. In fact, there is great overlap in the practices that bear these labels, and the labels themselves may be used interchangeably. The content of adult roles and relationships with students, as well as the desired outcome, varies considerably from program to program (see Flaxman et al. [1988] for a review). Within programs, the intensity of relationships may also vary. Freedman (1988) examined the quality of mentoring relationships in five programs that provided at-risk youth with "elder mentors." Three types of one-to-one relationships were observed: primary relationships, which were characterized by a high degree of attachment, trust, importance, and enjoyment; secondary relationships, which exhibited the same characteristics as those found in primary relationships, but in a less developed form; and nonsignificant relationships, pairings that were marked by distrust and distance.

Research on the effects of mentoring is scant. The available information suggests that mentoring can be a useful but *modest* approach for addressing students' needs. According to Flaxman et al.'s (1988) review of the literature, the goals for the relationship should be clear and within the mentor's power to achieve, and the mentor must be empathetic, able to assess accurately the needs of students, and able to apply resources appropriately and regularly. Research on the effectiveness of advocacy is also rare; the few studies that exist (see, e.g., Murray et al. 1981; Unger and Wandersman 1985) again suggest that this form of support works best when the adult role is structured around a few well-defined objectives that will in turn help the student to undertake specific actions.

However, few if any of the existing projects have been accompanied by quantitative evaluations that include carefully constructed comparison groups, statistical controls on initial student input differences, and statistical tests of effects on major student outcomes after a reasonable period of program operations (Flaxman et al. 1988; Smink 1990). To provide an empirical foundation for a discussion of programs that use adults from the community to assist the school success of at-risk youth, this report shows the results after two years of Project RAISE, a well-financed, multi-faceted program pursued by seven different community sponsors within the large urban school district of Baltimore.

RAISE Project Components and Samples

RAISE started in May 1988 with seven community sponsors that each made a seven-year commitment to provide support to groups of approximately 60 at-risk students, beginning from the time they enter grade six and following them through subsequent middle and high school grades. The sponsors include two churches (one predominantly black and one predominantly white), two universities (one predominantly black and one predominantly white), two large businesses (both predominantly white), and one fraternity (predominantly black). According to project materials, "the basic RAISE strategy is to create on a large scale the kind of sustained caring connections which can make a dramatic difference in the lives of very high risk children." RAISE expects to improve students' self-esteem and school-related behavior and progress, and to reduce high-risk behaviors such as substance abuse and teenage pregnancies.

Key components of the RAISE model include a full-time director and support staff who provide overall program development and administration for the set of seven sponsors; paid school-based advocates for each of the seven sponsors; and volunteer one-on-one mentors for each student served by the sponsors. The director and support staff are located at the Baltimore Mentoring Institute, a nonprofit agency created to manage a number of related activities in the city.

RAISE is funded by significant grants from two major local foundations and by annual contributions from the seven sponsoring organizations that together will total about $2 million over the seven-year project period. The RAISE combination of paid and volunteer components is intended to be at levels that could be widely replicated elsewhere if the project is proved successful.

The seven sponsors vary significantly in the degree to which they have implemented the RAISE model during the first two years of the project, which are summarized in table 1. The seven paid school-based advocates have worked from the outset with each sponsoring organization to serve RAISE students as "part counselor, part friend, and role model." The advocates' job includes monitoring attendance, grades, and behavior, building a relationship of trust with each student, and troubleshooting for individual students when necessary. All sponsors have recruited some volunteers to assist the advocates with after-school activities such as tutoring and recreation and with periodic events such

TABLE 1

Differences in RAISE Components by Sponsor

	RAISE COMPONENT				
SPONSOR	Advocate Established?	Mentors*	Subsample Selected?	Grade 5 Reading Score	Core Size
A	Yes	+	No	4.78	56
B	Yes	0	No	5.38	41
C	Yes	0	Yes	5.18	44
D	Yes	+	No	4.29	57
E	Yes	+ +	Yes	4.89	49
F	Yes	+ + +	Yes	5.73	43
G	Yes	+ + +	No	5.20	44

* Entries indicate that mentors have been provided for no students (zero), no more than about one-third of students (+), about one-half of students (+ +), or all students (+ + +).

as museum or zoo visits, attending athletic events, roller skating, or going to the movies. However, not all sponsors have established the one-to-one mentoring component of the RAISE model.

One-on-one mentoring is a particularly demanding component of RAISE that few of the sponsors have as yet been able to establish at a high level of implementation. The expected mentoring relationship with an individual student is one of sustained caring and attention by the adult volunteer. Although mentors may help with a student's academic or personal problems, the expected role is different from that of a tutor, professional counselor, or social worker (Flaxman et al. 1988). To achieve a strong supportive mentoring relationship that builds students' trust and provides effective role models for positive personal development by the students, RAISE expects a strong commitment of time and energy from the adult volunteers. The mentors must commit at least one year of weekly contacts that include biweekly face-to-face meetings. Mentors are provided with orientation and ongoing training by RAISE staff and are given regular information by the paid advocate about their students' programs and performance in school and elsewhere (Baltimore Mentoring Institute 1990).

One-on-one mentoring has been well established for each RAISE student by two sponsors (table 1, F and G), although one of these two took until the beginning of the second year of the project to achieve its high level of implementation. Three other RAISE sponsors have established one-on-one mentoring for some but not all student participants, ranging from matching about half of the students with mentors in one case (E), to having mentors for about one-third of students for the past half year in another case (A), to reaching less than ten students with active mentors in the third situation (D). Two sponsors do not have one-on-one mentors for any RAISE students but use their adult volunteers to assist with group activities or to work with different individual students on different occasions.

Another important way that some sponsors have deviated from the original RAISE design involves changes in the samples of students to be served. The original seven groups of students were identified for each sponsor by designating seven elementary schools in some of the city's most impoverished neighborhoods from which the students completing grade 5 in May 1988 would be eligible for RAISE. The initial design was for each sponsor to work with a group of about 60 students, which was thought to be an upper-limit case load for each school-based advocate and for sponsors to recruit and train volunteer mentors. However, the number of students eligible for RAISE from the designated feeder schools was often much greater than 60, and the sponsors coped with this greater number in different ways. Three sponsors accepted all the eligible students, even though their actual numbers of RAISE participants then ranged from 75 to 80. One sponsor who should have served 75 eligible students wound up with an actual RAISE sample of 67 because of apparent clerical errors in providing school lists, but no apparent biases were created by the sample reduction. Three other sponsors purposely eliminated about one-third of the students from their pools of eligible students (table 1, "Subsample Selected?"). Their original samples had ranged in size from 85 to 99, but subsamples were selected to achieve in each case a final group of about 60 students as actual RAISE participants. The process of selection in these three cases was not random, but based on which students showed most initial interest in the project, as reflected in two cases by providing home information of interest and in the third case by being active in the early months of the project. Our tabulations show that the eligible students who were eliminated by these three sponsors tended to have lower fifth-grade test scores and higher absence rates and were more likely to be male and to be designated as special education students. This nonrandom elimination of some eligible students introduces some complications to the original evaluation design.

The final actual samples for each sponsor differed in two other important ways that derived from initial differences of students and their elementary feeder schools. Average grade 5 reading scores for students differed by sponsor at the end of the school year when students first were selected for RAISE, with a difference of nearly one and one-half years in grade-level equivalent scores between the sponsors with the least well prepared and best-prepared students (D and F). Sponsors also differed in the degree to which their 60 students were dispersed across different middle schools, because of elementary school feeder patterns. Table 1 shows the number of students in the "core" middle school that included the largest number of students for each sponsor. Students in core schools should be easiest to serve because of their location at the site where the full-time paid advocate resides for each sponsor.

Each factor may have some influence on the effectiveness of each sponsor's RAISE programs over the first two years, which we will evaluate along a number of student outcome dimensions.

Evaluation of RAISE Effects

Because some RAISE sponsors eliminated students from the original pool of eligible participants, we adopted the following strategies to establish comparison points to evaluate possible effects of RAISE on student outcomes. First, we restricted our attention to comparisons between RAISE and non-RAISE students who are attending the same middle school, leaving the students who had been dropped by RAISE sponsors out of both comparison groups. Since each sponsor has a core middle school that most of their RAISE participants attend and where the paid full-time staff advocates are located, we focus on RAISE students who are most likely to receive the strongest assistance from the program. By omitting from the analyses all students who had been dropped from RAISE, we do not penalize the non-RAISE comparison groups with any potential negative bias of the individuals eliminated from RAISE. Thus, we are using students who began grade 6 at the same time in the same core middle schools to compare RAISE participants with other students who were not eligible for RAISE because their elementary feeder schools were not originally selected.

Second, we statistically control for a number of key student input variables with which students entered grade 6. These variables include grade 5 spring scores on standardized tests in reading, math, and language arts; student's sex and race; and student's age in grade 5, which indirectly indicates whether an individual had been left back one or more times in elementary school. Statistical controls are achieved through multiple regression analyses on selected student outcomes that include these input variables as well as a zero-one code for whether a student is enrolled in RAISE.

Our strategy of restricting attention to students in core middle schools will also hold constant many features of school policy and staffing not associated with RAISE, because RAISE and non-RAISE students in the same schools will have similar school programs outside RAISE. This initial strategy also aids in controlling for student inputs, because each core middle school draws from a defined geographic attendance area. All of our approaches, however, cannot *completely* discount possible bias in the analyses due to some RAISE sponsors' elimination of some student participants from the original eligible pool; thus we will also note any residual relationships with sponsor's selection practices.

Program Effects

Four sets of student outcomes are examined for possible effects of student participation in RAISE. Absence rate for 1989–90 (the second year of the RAISE project) is calculated by dividing the number of days absent by the number of days on the roll for each individual student.[1] Report card averages are calculated by averaging scores from the four quarterly marking periods of the 1989–90 school year, to establish English grades and overall grade point average. On-grade promotion rates are calculated from whether an individual student had been promoted at the end of grade 6 and at the end of grade 7. Student test score performance is measured by grade-equivalent scores on the California Achievement Tests (CAT) administered by the district to all students not in special education at the end of the spring 1990 term.

Table 2 summarizes the results of the overall RAISE project, combining the seven sponsors in a comparison of RAISE and non-RAISE students attending the same core middle schools with statistical controls for student inputs. Several statistics are provided for assessing the size and direction of possible RAISE effects. The unstandardized regression coefficient for an individual student's enrollment in RAISE, denoted by B, is produced by the multiple regression analyses that also include the student input measures. This B is thus an estimate of the difference in a selected student outcome due to participation in RAISE in a core middle school, controlling for student input differences. Our estimated "effect size," ES, is similar to the statistic often used in experimental studies and meta-analyses, and is calculated as the difference between RAISE and non-RAISE students (given by B) divided by the district standard deviation of the relevant outcome measure. The benchmark of .20 or greater for ES significance often used in educational research gives one standard for our assessments. We also provide a test of

TABLE 2

Comparison of RAISE and Non-RAISE Students in the Same Core Middle Schools on Selected Outcomes, with Controls on Student Inputs

Outcome	B	ES	*t*	α	RAISE Mean (*n*)	District Mean (*n*)
Absence rate 1989–90	−2.96	−.18	−3.04	.002***	18.38 (311)	16.77 (7,099)
Average English grade 1989–90	1.00	.14	2.36	.019**	71.72 (277)	73.05 (6,011)
Average math grade 1989–90	.33	.04	.62	.539	69.16 (261)	72.55 (5,845)
Grade point average 1989–90	.65	.09	1.60	.109	71.74 (274)	73.31 (6,982)
On-grade promotion 1988–90	.007	.02	.23	.816	.662 (314)	.759 (7,292)
Reading CAT GE 1990	−.087	−.04	−.09	.929	6.30 (264)	7.24 (5,817)
Math CAT GE 1990	−.127	−.06	−1.35	.176	7.01 (255)	7.72 (5,723)

NOTE.—Controls are sex, race, age at grade 5, special education designation, and grade 5 CAT scores in reading, math, and language. The analyses of the absence rate in 1989–90 add a control for the absence rate in grade 5.

** $P \leq .05$.

*** $P \leq .01$.

statistical significance for B, together with the probability (α) that the observed difference is due to chance. To help us judge how much RAISE effects reduce the actual risk level of students compared to a district standard, table 2 also presents the current mean of each outcome for RAISE students and for other students in the entire district who began grade six at the same time.

Table 2 gives evidence of both the potential and limitations of RAISE as it has been implemented in the first two years. We observe statistically significant positive effects on two outcomes—absence rates and report card grades. And although the positive effects are meaningful, the average RAISE student still has remaining attendance and grade performance problems that pose major risk factors for continued success in school and for completing high school without dropping out.

The effect on students' absence rate is shown in the first row of table 2. The reduction due to RAISE participation in annual absences of nearly 3 percent ($B = -2.96$) approaches the effect size that most analysts take seriously ($ES = -.18$) and attains a high level of statistical significance ($\alpha = .002$). When translated into the number of extra days of attendance in a 180-day school year, the estimated effect is that RAISE students will attend about one more week than comparable non-RAISE students in the same middle schools (2.96 percent × 180 = 5.3 additional days). While this increment in attendance is meaningful, table 2 shows it still leaves the average RAISE student with an annual absence rate of 18.38 percent, which is worse than the district average of 16.77 for students at this grade and remains a troublesome risk factor for success in later grades. Thus, our results indicate that RAISE taken as a whole has potential for improving student attendance, but the effects are not yet powerful enough to reduce the average absences to desirable rates. Since RAISE is not intended to upgrade the quality of schools where the students receive their classroom instruction, it remains to be seen whether RAISE alone can produce much greater gains in attendance and school performance.

The second and third rows of table 2 present findings of RAISE impacts on report card grades. Students enrolled in RAISE are getting better grades than other students in the same schools after controlling for student inputs, but these grades still remain below the district average. The positive RAISE effect is statistically significant for English grades but not for math grades. The positive RAISE effect for overall grade point average is in between the values for English and math and just misses achieving a minimum significance level of .10 (see table 2). But in terms of the effect size statistic, the RAISE effects are not large by conventional standards. Table 2 also shows that, even after the positive effects, the average RAISE student remains below district mean report card marks and close to the minimum passing grade of 70. Thus, participation in RAISE is shown to help students get better grades in English and perhaps in their overall grade point average than other comparable students in the same middle schools, but the impact is not strong enough to move the average RAISE student to a high level of classroom performance.

The final three rows of table 2 present results for student promotion rates and student achievement on standardized tests of reading and mathematics. We find no statistically significant differences between RAISE and non-RAISE students on any of these outcomes. However, the average RAISE student remains well below the district average on these outcomes and at levels that raise concerns about the chances of success in later grades, especially with regard to low promotion rates in middle grades, which are often precursors of dropping out at later grades. Table 2 shows that only about two-thirds of RAISE students have been promoted in each of the last years, compared to about three-quarters of students districtwide who began grade 6 at the same time (.662 vs. .759). So, even though RAISE has had some positive impacts on student attendance and grades, one of every three RAISE students was retained in grade at least once since the program began two years ago. Grade retention in elementary and middle grades has been shown to be a strong predictor of not completing high school (Shepard and Smith 1989), so many RAISE students remain in a high risk category for dropping out before high school graduation.

Sponsor Effects

A comparison of the seven different sponsors of RAISE programs is useful for further judgments on the potential and limitations of the first two years of operation, since we can consider the range of impacts across sponsors and whether it is related to known variations in the implementation of RAISE components or in the students served by each sponsor.

Table 3 summarizes the results for each individual RAISE sponsor on student absence rates and English grades, the two outcomes for which the overall positive impact of RAISE was most evident. Although the estimated sizes of effects are often much larger and more impressive in individual cases than the overall sizes reported in table 2, only three of seven sponsors report desirable effects that reach statistical significance on either of these two outcomes.

Sponsor C shows the largest differences between RAISE and non-RAISE students on both absence rates and English grades. In this

TABLE 3

Comparisons of 1989–90 Absence Rates and Average English Grades of RAISE and Non-RAISE Students in the Same Core Middle School, by Sponsor, with Controls on Student Inputs

	ABSENCE RATE				AVERAGE ENGLISH GRADE			
SPONSOR	B	ES	*t*	α	B	ES	*t*	α
A	1.46	.09	.54	.592	1.88	.26	1.58	.115
B	1.30	.08	.40	.687	−1.19	−.16	.83	.409
C	−7.51	−.45	7.51	.0004***	2.66	.37	2.40	.017**
D	−3.87	−.23	−1.36	.177	−1.07	−.15	1.68	.387
E	−4.84	−.29	−1.69	.094*	1.95	.27	1.68	.095*
F	−.78	−.05	−.30	.761	2.64	.37	2.34	.020**
G	−.60	−.04	−.15	.883	−.62	−.08	−.36	.722

* $P \leq .10$.
** $P \leq .05$.
*** $P \leq .01$.

case, the reduction in absence rates of 7.51 percent translates into 13 more days of school attendance, or over two weeks of added schooling for RAISE participants. The improvement in English grades is over two points on average for the RAISE students served by sponsor C, which is statistically significant with a very low probability of error. Sponsor E also shows sizable and statistically significant effects due to RAISE on both outcomes, though not quite as impressive as the results for sponsor C. For sponsor F, the estimated effects of RAISE are both in the desired direction, but only the result on English grades reaches an acceptable level of statistical significance. Thus, the overall positive effects observed for the combined sample of all RAISE sponsors was primarily produced by three sponsors who produced especially powerful desirable impacts on their RAISE participants.

RAISE Components and Effects

We can find several possible explanations for why some RAISE sponsors show stronger positive results by examining additional information about each sponsor's own RAISE program. Table 1 provides information for each sponsor on the degree to which one-on-one mentoring has been instituted (as an indicator of program implementation), the grade reading level of the average RAISE student (as a measure of student input), the number of RAISE students being served in the core middle school (as an indicator of the size and dispersion of student participants), and notes to indicate which sponsors eliminated students nonrandomly from the program and other special circumstances.

The use of one-on-one mentoring is not a consistent predictor of which RAISE sponsors produce the strongest positive results, although the set of three most effective RAISE sites includes two of the three sponsors with well-established mentoring components. Sponsors E and F are rated high on the implementation of mentoring and showed significant desirable impacts on their RAISE students. At the same time, the other effective sponsor (C) has the largest estimated RAISE impacts but did not use one-on-one mentoring at all. Clearly, other aspects of sponsor C's approach—such as the school-based advocate and the various activities between volunteer adults and RAISE students—were powerful enough to produce the observed benefits in student attendance and grades. Moreover, one sponsor (G) with a well-established mentoring component did not reflect significant average impacts in our analyses of differences between relevant RAISE and non-RAISE students. But sponsor G is a special situation among RAISE participants, in that students did not shift from an elementary school setting to a middle school environment until grade 7, which might explain difficulties in this case of producing RAISE effects at the end of year 2. The other three sponsors that failed to show consistent significant effects on student outcomes in our analyses did not have strong mentoring components, pairing only a very small number of individual students with mentors (sponsors A and D) or having no one-on-one mentoring for any students (sponsor B). Thus, although we find one exceptional sponsor with strong effects and no mentors and one exceptional sponsor with well-established mentors and no consistent effects, the results for the remaining five sponsors support a conclusion that one-on-one mentoring is an important RAISE component: three with weak mentoring showed no effects and two with strong mentoring showed significant positive effects. It seems reasonable to conclude that, although strong one-on-one mentoring is not an *essential* component of an effective program that uses outside adults to assist at-risk middle school students, the RAISE model is much more likely to show positive effects when one-on-one mentoring *has* been strongly implemented.

Other aspects of RAISE implementation also remain as possible explanations for differential sponsor effectiveness, including the reading scores with which students enter the programs of the different sponsors and the reduction of the numbers of students to be served by certain sponsors. The association of student input differences with sponsor effectiveness is only suggestive, if the grade 5 reading scores and core sizes shown in table 3 are used as points of comparison. The two sponsors with the lowest average student input reading scores and with the largest number of students to be served in the core school (A and D) are also the two that show no consistent effects across any of the outcomes we examined and whose students remain most at risk in their further schooling. But, except for these two least effective sponsors, there is no other pattern between student inputs and sponsor effectiveness. While this evidence only suggests that RAISE effectiveness is more likely for smaller groups of students who do not begin far below grade level, other findings are also consistent with this conclusion.

Table 3 also shows that the three most effective sponsors are the ones that had selected a nonrandom sample of actual RAISE participants from the original pool of eligible students. Although we cannot completely rule out the possibility that these positive results were due to sample selectivity rather than actual RAISE program impacts, additional analyses support a more general interpretation: RAISE effects are more likely for sponsors who serve at-risk students with less severe initial educational disadvantages. When we repeated the analyses shown in tables 2 and 3 on the original samples that included all eligible students as RAISE participants, we found similar results. Adding all eligible students to the analysis did not substantially alter the effects of RAISE, even though only some of them participated in the program of three sponsors. As might be expected, the results were slightly smaller in size in these reanalyses, but these effects remained statistically significant in the same patterns reported in tables 1 and 2.

Some practical implications are suggested when we combine our conclusion that sponsors who reduced their original samples had the largest nonspurious RAISE impacts with our previous observation that the two sponsors who began with the largest and most disadvantaged student groups showed no effects. It appears that the size and composition of student groups to be served require major differences in resources if programs such as RAISE are to be successful. Programs that begin with student groups that are very large or greatly behind academically will have a much greater struggle to demonstrate positive effects.

Further Implications for Research and Practice

This evaluation is one of the first that uses comparison groups and statistical tests to judge the effects of a well-financed program using adult advocates or mentors. It provides some additional new perspectives on major practical and research questions.

Both the potential and limitations of programs such as RAISE emerge from the evaluation results after two years of operations. We find it possible to help students make impressive gains in school attendance and in report card marks, but the average gains after two years are not sufficient to eliminate the academic risks with which students entered the program. Even after the RAISE benefits, the average student continued to have serious problems of absenteeism and low grades compared with the typical student in the district. And RAISE has not yet had measurable positive impacts on student standardized test scores or promotion rates in the middle grades. Nevertheless, the student behaviors where RAISE has been successful in its first two years can be viewed as steps in a sequence to improve students' academic chances as the program continues.

School attendance, on which RAISE demonstrated a positive impact, is a behavior that is most open to short-term improvements and that can lead to advances in other school outcomes. Good attendance may be more completely under the control of individual students and more susceptible to positive influences by adult advocates or mentors than other school behaviors. In contrast to report card grades and promotion rates, for which teachers make the major decisions, every student can have good attendance. To be sure, some teachers can induce better student attendance through more engaging lessons and more positive relations with students. But any student with absenteeism problems can improve daily attendance with extra effort and effective support from family and meaningful adults. Effective encouragement by mentors or advocates to improve student attendance may require less training and program management than other support activities such as academic tutoring or negotiating with teachers or school officials on behalf of students. Thus, attendance rates seem to be a student outcome on which adult mentor and advocacy programs can focus to be effective in the short run.

Good school attendance can often be a building block to other student behaviors required for school success. Students' attendance rates are often closely tied to their school report card grades, because many teachers will mark down students who have higher absenteeism levels than the rest of the class and some teachers will automatically fail students who have missed a significant portion of the term. Course failures due to poor attendance can lead to higher retention rates. Student learning as measured by standardized test scores can also be expected to suffer as a result of poor attendance, because absent students will miss instruction and engage in less drill and practice in the basic skills covered by tests. So a program of assistance by outside adults that focuses on improving student attendance may have a cumulative effect over time on other academic outcomes.

Our findings that RAISE affects report card grades may be partially explained by the improvements in attendance that may lead to better grades. We found strong RAISE effects only for English grades, so it is likely that direct support for academic learning by RAISE adult advocates and volunteers was also responsible, through activities such as assistance with completing homework, tutoring in basic skill areas, or assistance in learning activities such as reading practice. More effective assistance with academic learning may require more specialized training of RAISE adults or more coordination with the ongoing school instructional program, but our evaluation of RAISE suggests that direct academic activities can be a successful early part of programs using outside adult advocates or mentors.

Major issues also emerge from our evaluation of RAISE about the successful implementation and coordination of program components. Programs that seek to make one-on-one mentoring a key component face several issues. Recruiting mentors and having them sustain successful relationships with at-risk students has been a major challenge for most RAISE sponsors. Some sponsoring organizations (such as the two participating churches) were able from the beginning to locate sufficient numbers of adults who committed to the mentor role and responsibilities. But most sponsors have taken more time to establish mentoring activities and are still working to recruit an individual mentor for every RAISE student. In some cases, sponsors have followed a process in which sets of adult volunteers engage in group activities with groups of RAISE students to assist in identifying one-on-one pairs that seem to be good matches for a sustained positive relationship. But when the requirements for mentors are major and regular commitments of time and energy, it will often take a long time before a sponsoring organization can build a cadre of committed mentors. During this time, other useful activities using adult volunteers in less demanding roles can be established by sponsors as part of their process of establishing and supplementing an effective mentoring component.

The mentor's role is usually described as being a caring adult to support a student's efforts to succeed at major goals, but this conception raises questions of implementation and coordination with paid adult advocates in the program and with other adults in the student's school and home. Mentors are usually not intended to assume the supervisory and disciplinary roles of parents and teachers, but to provide a positive uncriticizing reference in the student's life. Yet, some of the student behaviors that appear most responsive to influence by outside adults, such as improved school attendance, may require adult monitoring and pressure that goes beyond the theoretical role of mentors or beyond the understandable preferences of some adults who actually fill these roles to avoid possible confrontations with their students. This raises the question of whether others such as the paid adult advocate might better handle the supervisory and disciplinary activities while mentors continued to focus on positive supports and incentives in their relationships with individual students, or whether both the definition and training of mentors should address the needs for both adult support and constructive criticism.

Further evaluations of RAISE or other programs with similar components and goals are needed to learn whether effects become stronger as a program continues beyond its formative years and whether certain students respond better and benefit more from adult advocates or mentors as they move through the middle and secondary grades. RAISE managers are using our evaluation of the project's first two years to intensify and focus their efforts for the future. They expect one-on-one mentoring to gradually become available for most student participants in all but one sponsoring organization, and they intend to concentrate more on improving student school attendance and on working out arrangements with school officials to minimize retentions. Closer attention will be paid by both RAISE practitioners and evaluators to which students participate most in each type of RAISE activities to learn about individual differences in student responsiveness to RAISE offerings and to identify the RAISE components and activities that have the most impact on particular student outcomes.

Notes

This research was supported by The Abell Foundation, Baltimore, Maryland. The authors gratefully acknowledge the support and assistance of Kalman Hettleman, Director of the Baltimore Mentoring Institute, Richard Rowe, Director of Project RAISE I, and A. C. Hubbard, President of the Board of Trustees of RAISE, Inc. Opinions expressed herein are our own, and no endorsement from the sponsoring and supporting organizations should be inferred.

1. To avoid problems sometimes associated with extreme values on ratio scales, we eliminated students who either (*a*) missed more than 135 days of the 180-day school year or (*b*) were on the roll for less than 45 days. This truncation also obviates problems due to schools that may retain names on the roll that likely have left the district, in order to gain an enrollment advantage for staffing calculations. We repeated all analyses using untruncated absence rates and observed no substantial changes in reported results.

References

Baltimore Mentoring Institute. *RAISE Fact Sheet.* Baltimore: Baltimore Mentoring Institute, 1990.

Berger, J. "East Harlem Students Clutch a College Dream." *New York Times,* August 27, 1989.

Freedman, M. *Partners in Growth: Elder Mentors and At-Risk Youth.* Philadelphia: Public/Private Ventures, 1988.

Flaxman, E., C. Ascher, and C. Harrington. *Mentoring Programs and Practices: An Analysis of the Literature.* New York: Teachers College Institute for Urban and Minority Education, 1988.

Murray, C. A., B. Bourgue, and S. Mileff. *The National Evaluation of the Cities in Schools Program, Report no. 4: Final Report.* Washington, D.C.: American Institutes for Research, 1981.

New York State Mentoring Committee. *Directory of Mentoring Programs in New York State.* New York: Governor's School & Business Alliance, 1989.

One to One. *One to One: Mobilizing for the Next Generation.* Washington, D.C.: One to One, 1990.

Shepard, L. A., and M. L. Smith. *Flunking Grades: Research and Policies on Retention.* Philadelphia: Falmer, 1989.

Smink, J. *Mentoring Programs for At-Risk Youth.* Clemson, S.C.: National Dropout Prevention Program, 1990.

August 1991

Educational Evaluation and Policy Analysis
Winter, 1991, Vol. 13, No. 4, pp. 353–362

Lessons From the Field: Case Studies of Evolving Schoolwide Projects

Linda F. Winfield
Center for Research on Effective Schooling for Disadvantaged Students
The Johns Hopkins University

This study describes changes that occurred in one of the nation's largest urban school systems on the East Coast following passage of the Hawkins-Stafford Amendments. Case study methods were used to describe the central office and system role and changes at the elementary school level in selected sites. A major emphasis of central office framework for schoolwide projects (SWPs) was school-based management and instructional frameworks based on effective schools research. The primary type of instructional intervention at the school level was reduction of class size during reading and math instruction. Schoolwide projects offer the potential for improving learning outcomes of disadvantaged students but require coordinated and direct support from the central office and district.

The promise and potential of ESEA (Elementary and Secondary Education Act) Chapter I for improving the school achievement of economically disadvantaged students was part of the underlying rationale for its creation. This focus also contributes to its continued success among policymakers and practitioners as a categorical funding program. Both of these groups recognize the need to devote additional resources to assist schools serving student populations where high levels of poverty have a negative impact on schooling conditions and learning. Evaluations of Chapter I (formerly Title I) programs have been mixed but have generally failed to find substantial long-term achievement effects for students receiving services (Carter, 1984). The variability of program effects, while due in part to methodological differences, is also due to the variation in the actual educational program implemented. Chapter I is a funding program that provides supplemental services to the regular school program. The typical mode of delivery of instructional services has been the "pullout." Previous research has documented the disruptive impact of pullouts, the waste of materials and time in trying to keep noneligible children from benefiting from Chapter I services, and the limitations on use of effective programs imposed by the principle that only test-eligible children may be served (Allington & Johnston, 1989; Glass & Smith, 1977; Leinhardt, Bickel, & Palley, 1982; Winfield, 1986a). Moreover, in many of these schools belief systems develop among teachers and administrators in which they abdicate the responsibility for improving the learning of students receiving Chapter I services (Winfield, 1986b).

Nearly over a decade ago, case studies of schools in low-income urban and rural areas revealed that schools which "succeeded beyond expectations in teaching reading" (Venezky & Winfield, 1979) were those in which the principal and staff had made a conscious decision to improve the achievement of all students and had targeted high school-wide achievement as a goal. Principals in these schools included Chapter 1 (then Title 1) as part of the overall strategy. Although compliance with federal regulations provided cer-

Copyright 1991 by the American Educational Research Association. Reprinted by permission of the publisher.

tain restrictions, these schools operated a "student-centered" delivery of instructional services. This meant that all available building-level instructional resources (Chapter I, district, or otherwise) were coordinated and targeted to support and reinforce student learning in the core instructional program rather than the more typical case where each supplementary program had its own methods, materials, philosophies, and approaches. This latter scenario resulted in a fragmented instructional program for low-achieving students.

The Hawkins-Stafford Amendments (1988) which allow the use of Chapter I funding for schoolwide projects (SWPs) in schools where 75% or more of the students are economically disadvantaged are designed to reduce the fragmentation and to upgrade the entire school program. The flexibility in federal regulations comes at a time when the knowledge base has been advanced concerning effective schools (Purkey & Smith, 1983), the change process (Fullan, 1982), and successful programs in urban schools (Slavin, Madden, & Karweit, 1989). A major task confronting urban school systems and schools is how to make use of this new knowledge and also take advantage of the increased flexibility to improve the learning outcomes of low-achieving students. These opportunities come at a time when poverty has increased dramatically in major urban school districts (Wacquant & Wilson, 1989) and when contextual factors, such as size, demographics, diversity, density, a growing "underclass," the underground economy of drugs, the politics of school boards, and an eroding tax rate create uncertainty and turbulence in the school environment (Englert, 1989).

Urban School Systems

Some researchers suggest that in response to uncertainty, schools and districts develop large, complex, bureaucracies (Bidwell, 1965) that are characterized by rigidity and a variety of dysfunctions (Levine, 1978). In large, urban districts, the response is characterized by a tendency toward disengagement from instruction (Meyer & Rowan, 1978). District and central office administrators are generally removed from what goes on in classrooms. In general, the governance and control of many schools serving disadvantaged students are often fragmented by competing groups (e.g., unions, school boards, state departments, and special interest groups) and programs (e.g., Chapter I, migrant, special education, bilingual, curriculum and instruction, budget, personnel, and school operations). There is a high degree of role differentiation and specialization at the central level, and individuals and groups become territorial regarding their expertise, budget, and their constituencies. Services provided to the schools are seldom coordinated, and school building principals must deal with four or five central staff persons for a simple request. Central office administrators and policies at this level influence coordination of efforts and instructional collaboration at the school level (Birman, 1981; Kimbrough & Hill, 1981). Thus, the purpose and intent of schoolwide projects which focus on upgrading the whole school program may be difficult for schools to attain, given the competing demands for central office groups. In general, when school districts reorganize central administration, the realized improvements are marginal and confounded with other simultaneous changes (March, 1978). The tinkering with duties and position titles rarely has an impact on the instructional process in classrooms. The difference between those districts in which schools successfully change and the typical school is the concept of "connectedness" (Wimpleberg, 1989). Wimpleberg (1989) notes that it is unlikely that schools will act on their own to improve or that school systems will have resources to employ the needed specialists to assist in the process of change. That type of technical assistance must be provided from personnel rather than through paperwork (Eubanks & Levine, 1987). Other studies, however, have identified examples of strong leadership by large urban school superintendents who shape district-wide conditions for improving schools within a context of broad community support (Hill, Wise, & Shapiro, 1989).

The purpose of the case study was to describe changes in a major urban school system and schoolwide project schools following the passage of the Hawkins-Stafford Amendments. Between July 1989 and July 1990, central and district office SWP meetings and staff development sessions were attended, and one- to two-day site visits were conducted in 11 schoolwide project schools. School system documents, school SWP proposals, and other reports were examined. A more thorough description of this study can be found in Winfield and Stringfield (1990).

School System Context

The Chapter I program in a major urban school system serves 162 schools and receives approximately $50 million in Chapter I funds annually. Since 1983, various initiatives targeted toward improving the achievement of Chapter I schools had been initiated by the superintendent, who can be described as a demanding instructional leader. She manages a $1 billion budget yet places the education of the 250,000 children ahead of everything else. One of the past initiatives targeted the improvement of 26 Chapter I schools over a three-year period beginning in 1983. Funds from a private foundation and Chapter I funds were used to support a school-based planning and implementation process. As the third year of the project began, the central office felt that in some schools additional human resources were needed to change the historical patterns of low student achievement which existed. Thus, for example, teachers were hired to staff full-day kindergartens, and a permanent substitute was assigned to the schools. For the 1986–1987 year, the superintendent opted to designate 11 of these schools previously targeted as schoolwide projects and to pay the matching share then required for noneligible students who were receiving services. When the Chapter I guidelines were changed in 1988, the school system expanded the number in the program rapidly. Currently, approximately half of all of the elementary schools in this system are SWPs.

At the same time these initiatives were under way, a system-wide Chapter I Task Force that had been meeting since 1987 made a recommendation based on student outcome data to expand SWPs. The task force consisted of all of the major special interest groups and stakeholders (e.g., central office staff from budget, special education, curriculum, and compensatory programs, as well as district superintendents, teachers, and principals in Chapter I schools). The major task was to develop a comprehensive compensatory program, designed to improve student achievement which would be phased in over a two-year period. One former task force member interviewed said: "It was a working group . . . it brought everybody to the table. . . . we didn't always agree but we knew that something had to be done to improve . . . that's the bottom line."

The school systems approach to SWP identifies five main thrusts: (a) a whole-school approach which supports student success in the daily program, provides special support for students who require it, and is based on the "effective schools" research; (b) school-based management which requires that the school staff and parents determine the nature of the intervention within specified program guidelines and contractual requirements. Chapter I funds are provided to each school as a block grant (total averaging about $250,000–$300,000 or $900/pupil). (c) Monitoring individual student, class, and school performance on an ongoing basis and giving particular attention to those students targeted for intensive services and those who would be designated as Chapter I eligible should they attend a nonschoolwide project; (d) district-based support provided by the central and subdistrict offices to provide parent and staff training on an "as requested" basis. This support targeted leadership development and team building, ongoing leadership team meetings for principals and key staff, and monitoring school improvement plans. (e) Concentration of resources which indicates that funds beyond the minimum amounts would be committed from Chapter 1 and operating budgets.

As part of its commitment to SWPs, the central office designated an office of schoolwide projects with a director and manager to develop and oversee the implementation. Mr. D., the selected manager, had had no prior central office experience but had been a highly successful elementary principal for some years and was credited with "turning around" the dismally low performance in an extremely impoverished urban school. Mr. D. indicated:

> From my own experience to change what's happening in schools, staff development for principals is critical because most of them don't know what to do . . . teachers have to be supported because many of them are scared to change, and direct services to the school have to be expanded but also coordinated.

Throughout his tenure, he developed the operational guidelines for implementing SWP but was also the role model for principals and the chief advocate for students, teachers, and principals in schoolwide projects. On many occasions, he indicated that schools needed to undergo an "awareness and orientation phase" that primes them for changes in how they traditionally deliver services to Chapter I eligible students and how to participate in decision making on a schoolwide level.

System-Level Support for School Change

According to Mr. D. and other principals in SWP sites interviewed, school-based planning and site-based management are processes which are not easy to carry out effectively. Thus, principals and teachers require continual coaching, encouraging, admonishing, recognition, and incentives in order to get them to "buy into" the process and to implement a schoolwide intervention. Principals and teachers have traditionally selected instructional materials and made decisions about a particular program or focus; however, few in SWP sites had been involved with making decisions which affected the whole school or with reaching a consensus concerning decisions, such as hiring an additional math resource teacher or eliminating the reading laboratory.

There were newly created positions of specialized SWP personnel to assist schools with these decisions in their leadership team meetings. These persons were knowledgeable about change and about the instructional process, and school-based management provided the services of internal change agents. A position titled "Instructional Interventionist" served as a liaison between the subdistrict and the SWP site. The interventionists were action-oriented and participated in principal-led monthly leadership team meetings in each SWP school, and they organized ongoing staff development and cross-school sharing for principals and staff. In addition, they coordinated, directed, and provided staff development for instructional support teachers; provided assistance to the principal in arriving at a workable school improvement plan; and ensured that all materials and supplies purchased were related to the school's detailed instructional improvement plans.

The instructional support teachers (ISTs) held teacher-level positions in each subdistrict; each IST was responsible for overseeing two SWPs and worked directly with the principal and school personnel. They were in each school two to three days a week, depending on needs at the individual SWP site. They served as a "trouble shooter" and an implementation coach for the principal. In addition, they provided staff development and worked with their school-based counterparts, the program support teacher (PST). These teacher-level positions were based at the school and were selected from the school staff by the principal. The PSTs were considered by their peers as a "master" or "mentor" teacher. They provided instruction to students for 90 minutes a day and spent the remainder of the time working directly with the principal, new teachers, and other staff in implementing the schoolwide plan. They monitored student progress, participated in leadership team meetings, and did demonstration lessons. Interviews with these staff indicated that their positions were "labor intensive, demanding and required long hours," yet they were highly sought after as evidenced by the number of teachers applying and taking the oral and written test. In interviews regular classroom teachers indicated that the positions were "higher status" and meant to them "master teacher" even though PSTs and ISTs received the same salary as a regular teacher. The system of specialized positions was dynamic—one in which new talent was constantly sought as veteran program support teachers were promoted into IST positions, and ISTs moved up to instructional interventionists. Several of the instructional interventionists became building principals.

Instructional Frameworks

Another major objective of implementing SWPs was to change from a traditional "pull-out" model to a whole-school instructional focus. The central office developed and provided staff development in four instructional frameworks from which schools could choose in order. The frameworks were general and included factors such as high expectations, monitoring, positive school climate, and team work. They also included, however, classroom-based strategies such as cooperative learning, active teaching and learning, and effective lessons. Attendance at staff development was highly encouraged, and teachers were paid; however, it was voluntary. Thus, the use of the diverse frameworks varied across school sites. In some sites, the framework was not a salient part of the observed school program. In other sites, the framework provided a common language for staff to use in discussing students and instructional matters, or it served to facilitate team building. In these settings, the frameworks helped to create a sense of community among staff, allowing them to coalesce around common goals. This was particularly evident during staff development sessions when principals and teachers from all SWP sites were divided into groups on the basis of the framework adopted. The particular instructional framework itself was not as important as allowing principals and teachers to select and adapt one which they felt was most appropriate to their school.

Parent Involvement

Another concern in SWP sites was to involve parents in the educational process of their children. Each school's SWP proposal was required to delineate ways in which the site would conduct parent involvement activities. Schools were also required to include in their budgets funding for a school community coordinator. This individual initiated strategies to improve attendance. At several sites, he or she was responsible for implementing a daily system of identifying all absent students in order to make immediate contact with the home. He or she also coordinated and directed parent workshops over the school year. "Parent scholars"—parents from the community who assisted in the classrooms—were also funded out of SWP budgets. These assistants were provided with a modest stipend and worked in 10-week cycles. Parents were observed assisting in classrooms, the library, computer lab, and lunchroom. Each SWP site was provided with a parent trainer, who visited the site regularly to assist in recruiting and training community assistants and to assist in other parent involvement activities. Each site also had a trained home demonstrator, whose sole purpose was to make home visits and to work directly with parents on learning readiness, on helping their child with homework, and on other school-related activities. These personnel provided systematic support and increased the number of parents involved in SWP school activities.

What Are the Major Types of Interventions?

The 11 SWP sites in the sample used their Chapter 1 funds in a variety of ways. In the first years, some of the funds were used to purchase needed materials such as science kits, math manipulatives, and classroom literature libraries. One site used funds to extend the school year by 22 days. Nearly all of the schools established an additional teaching position to lower the teacher–student ratio during math and reading instruction. Approximately half reduced class size in classes with the lowest achieving students. In over half of the SWP sites, the additional teaching position eliminated split grade classes. In some schools, the program support teacher, SWP reading and/or math resource specialist, and basic skills teacher provided the entire lesson to the whole class on a scheduled basis. Other schools developed team teaching models in which SWP personnel taught in the classroom with the regular teacher.

What Does a SWP Look Like?

School A

Constructed in 1937, School A is a small school building in the middle of a neighborhood that is rapidly undergoing regentrification. A few blocks to the east are new town homes and renovated row houses selling for $250,000 and up. A few blocks to the west are the remnants of three high-rise-project buildings that are slated for demolition. Because of population shifts, the school enrollment has declined from 700 to 397. Ninety percent of the students in the school are eligible for free lunch. Although the school population is predominantly Black (81%), there are also Asian, (3%), Hispanic, (4%) and White (12%) students attending the school. The school enrolls children from kindergarten to eighth grade and has 13 regular grade classes and 10 special education classes. Moderately and severely handicapped children are bused from outside the immediate school neighborhood. The staff have devised activities to integrate many of these students into as much of the school day as possible.

Although it is not officially a magnet school, it enjoys a "good reputation," and according to the current principal, parents are "clamoring" to get their children enrolled. This school finished among the top 10 in science and mathematics in the district and has had a full-time science room with a science teacher for only the last four years. As one enters the school, it becomes immediately apparent from the school banner, displays of students' work, trophies, and awards that someone has fostered and maintained a sense of school pride and spirit. In the school office, a two-page handout for substitute teachers is noticeably displayed and provides essential information on lesson plans, roll book, homework, lines, classroom management, school procedures, and academic notes on each subject area. The first line reads: "We are a school-wide project school."

Mr. A., the principal, is completing his first year at School A, but he has been a principal in urban schools for the past 22 years. He credits the success of the SWP to the former principal, who had created the basic idea. He felt that the plan was "teacher intensive," indicating that the funds were used primarily for teachers to reduce class size in reading, and he said, "We're not lacking for materials, but most of the money was spent on personnel. In the lower grade, two teachers assist in the primary grade reading cycle. One works with the upper grade reading/language arts." He attributes the successful implementation of the plan to the ongoing staff development and an active pupil support committee, which meets twice a month to discuss alternative interventions for individual students having academic problems. He said, "Having paid staff development and meeting time for pupil support committee meetings has been a great advantage." He described the staff as "stable and very strong-willed but capable and very caring" of the youngsters that they taught.

Teachers interviewed indicated that since becoming a SWP, the biggest change was that all of the faculty provided input into the plan. Other teachers indicated the increased flexibility as an advantage. One teacher said, "I never liked the pullout model . . . . the coordination makes sense . . . they're not freight packages—they're children." Another teacher noted: "Before, the classroom assistants could only teach certain students; now they can deal with all of the kids." The teacher who also chairs the "Climate" committee for the school credited the paid meeting time for the pupil support committee, which he indicates "allows us to be more systematic in finding out and doing something with kids who are having problems. . . . These teachers here really care . . . this place is like one big family, and we support each other."

According to the program support teacher, Mrs. Bee, who had been at the school for 22 years, the "thinking skills" was an area that staff decided to work on even prior to becoming a SWP. Their decision to adopt the framework on thinking skills, she indicates, was an easy one. Mrs. Bee explained that although she is "teacher of record" for a group of 12 kindergarteners that are reading, and 15 of the lowest achieving first graders, the majority of her time is spent in various classrooms conducting demonstration lessons or co-teaching with teachers. She explains that "teacher of record" is a SWP concept meaning that the person who provides the instruction in reading, for example, is also responsible for monitoring and improving student progress and grading. Her other responsibilities include assisting classroom teachers. She says:

> If a teacher is absent, I'll go in during the reading period so that the reading instruction is not disrupted. We have a new teacher in the school, and I was in her class conducting the reading lesson for the first month or so . . . since also, along with Mrs. G, the other reading teacher, I give an informal reading assessment to all students 3 times a year so that some don't fall through the cracks.

She viewed an important aspect of her position as monitoring and said that SWP requires a lot of paper work; however, she indicated, "teachers have all the information on individual students in one place, the SWP record book—grades, end-of-unit tests, citywide test, teacher-made tests, homework assignments. I collect these every 6 weeks from each teacher and review them."

The school's focus on integration of special education students, as well as commitment to teach all students, was observed during the reading cycle. An excerpt from the observation follows:

> Mrs. Bee began her routine that apparently all of the children know. It is a song with hand motions that the children do that captures their attention and tells them to put their thinking caps on. She introduces the lesson by saying, "Today we are going to talk about beginning sounds," and draws a picture of a hat on the board and writes —at. Now what belong in this space? The children respond "H" and she writes it in. She continues to introduce word families and sounds that will later be used in a "big book" story she reads. She is animated, moves around the room, calling on the whole group and individual children, including the special ed students, to respond to provide the beginning sounds for the pictures and word families on the board.

One would not have known that the students in this classroom were classified as "trainable mentally retarded" except for the size and age of the youngsters seated in the back and one little girl who imitated the behavior of the other children but couldn't understand. Still she raised her hand to respond to questions and tried to write the letters which upon observation were unintelligible scribbles. Mrs. Bee praised her for trying and later told me that "she really tries, but her problems are too severe."

School B

School B, built in the 1970s, is a factory-like structure that takes up nearly a city block and overshadows the small rowhouses in its immediate neighborhood. The school is surrounded on one side by a large outdoor play area, and in the front by a mixture of both well-kept and decrepit boarded-up row houses. An influx of young families with small children in the area has caused an increase in the school's enrollment. At the time of the visit, the school was past the capacity (900) of the building. Seventy-one percent of students are Hispanic, 21% White, and 7% Black. Every available space at the school is filled. One kindergarten class and one class of third graders use rented space in an adjacent church building. The science, art, and music rooms have been converted to classrooms. Specialist teachers go from classroom to classroom. The small conference room also functions as a lunch room for teachers. One is immediately overwhelmed by the sheer size of the building with its huge hallways and extremely high ceilings with exposed pipes. Despite the massive number of students, entrance, dismissal, and a fire drill were experienced without chaos and were quite orderly. Prominently displayed in the lobby is the "Creating Success" logo, the instructional framework chosen by the school.

According to the school community coordinator, a young bilingual Hispanic female who has lived in the neighborhood for 15–20 years, the big change in the neighborhood occurred about 5–6 years ago. "Many stable families who could afford to move left the area . . . we have a lot of young families with many children—some who have just come over from Puerto Rico . . . we also lose kids whose families return." In addition to her home visits requested by teachers and parent workshops, a substantial amount of time is spent referring parents to community agencies, taking parents for appointments, translating for parents who don't speak English, providing clothes and emergency shelters, and interpreting report card marks. She had served as home and school president prior to becoming a school community coordinator.

Mr C., the principal, is a high-energy, fast-paced, organized, and task-oriented individual. He stands in the hall and greets each child by name, handing out small rewards for good behavior, for perfect attendance, and for reading. He indicated that the school as a whole is challenged to read one million pages between September and June. Since becoming á SWP, he felt that because he was allowed more flexibility in how he could use his staff he stressed co-teaching models. In the school, SWP teachers team taught with grade teachers to reduce the student–teacher ratio during language arts periods. The math resource teacher and assistant teamed with three grade teachers to reduce the student–teacher ratio for math instruction for one hour each day. The ESOL teacher worked in the classroom with ESOL students. Funds also paid for planning time for the ESOL teacher to plan with grade teachers. Mr. C.'s background is in reading/language arts, and he teaches during one class period each day.

> He states: "The only way children learn to read is through reading . . . through frequent, positive interactions with a variety of meaningful texts. They learn to construct meaning. . . . The developmental process is supported by systematic, explicit instruction in phonics/word attack skills. We increased the amount of time from 1 hour to 90 minutes, emphasizing literature-based instruction and thematic unit planning. Because of SWP, we reduce class size and student–teacher ratio during reading/language arts instruction. We also have on-site staff development in implementing [a] literature-based reading program provided by a local university.

Teachers receive 3 graduate level credits. He indicates that this has been very successful in helping teachers to learn how to implement literature-based instruction.

Other teachers interviewed noted other benefits of being a SWP. One teacher indicated: "We are a big school, and we have a lot of funds poured into us, but we were told this is how you have to spend it regardless of whether students needed it or not." Another said: "Now we're able to get more personnel. . . . We used to have classes ability grouped, and some classes were not Chapter 1 eligible but were still in need of additional help. Before, I couldn't serve them." Other teachers noted that paid meeting times, on-site training, and availability of funds to purchase sets of literature books and expand classroom libraries were important benefits of being a SWP.

Conclusions

From the brief descriptions provided here, one can note that SWP schools in the process of change look much like other schools that are making conscious attempts to improve classroom instruction and improve upon existing programs. Because of the high concentration of poverty, many of the schools are plagued by staff vacancies, operating budget cuts due to declining enrollments, and high student mobility. However, with the assistance provided by SWP personnel, the schools are grappling with issues, such as how to make schoolwide decisions, how to create effective working plans for improvement, how to integrate other existing categorical programs into a coherent instructional program, how to allocate the available resources effectively, how to provide on-going support to classroom teachers, and how to deliver higher quality instruction to disadvantaged students.

In order for schools to change from a traditional Chapter I program to a more integrated focus on all students, parallel changes must be made at the central office. Not only must these offices become more "connected" to the instructional process, but they must also be organized in such a way as to provide effective coordination and delivery of direct service to schools becoming SWPs. School systems will also have to invest heavily in human resources and professional development at all levels. High-poverty schools, such as SWP sites, have tremendous needs for direct, on-site, systematic assistance in changing existing structures and negative belief systems; more intensive professional development in collaborative teaching models and subject-matter instruction; more proven high-quality educational interventions for students experiencing academic difficulties; and strong reciprocal agreements with teacher training programs to aid in recruitment and development. Schoolwide projects have the potential for improving the learning outcomes of large numbers of disadvantaged students. However, this potential will be met only if adequate support for change is provided at the central or district level and if sufficient resources are devoted to human resources and professional development.

Note

A version of this paper was presented at the National Conference on Educating Black Children, June 2, 1990, Los Angeles, California.

References

Allington, R. L., & Johnston, P. (1989). Coordination, collaboration, and consistency: The redesign of compensatory and special education interventions. In R. E. Slavin, N. L. Karweit, & N. A. Madden (Eds.), *Effective programs for students at risk* (pp. 320–354). Boston: Allyn & Bacon.

Bidwell, C. E. (1965). The school as a formal organization. In J. G. March (Ed.), *Handbook of organizations* (pp. 972–1022). Chicago: Rand McNally.

Birman, B. F. (1981). Problems of overlap between Title I and P. L. 94–142: Implications for the federal role in education. *Educational Evaluation and Policy Analysis, 3*, 5–9.

Carter, L. (1984). The sustaining effects study of compensatory and elementary education. *Educational Researcher, 13*(7), 4–13.

Englert, R. M. (1989, October). *Understanding the urban context of schools in large cities*. Paper presented at the 1989 convention of the University Council for Educational Administration, Paradise Valley, AZ.

Eubanks, E. E., & Levine, D. U. (1987). Administrative and organizational arrangements and considerations in the effective schools movement. In D. S. Strickland & E. J. Cooper (Eds.), *Educating black children: America's challenge* (pp. 19–32). Washington, DC: Bureau of Educational Research, School of Education, Howard University.

Fullan, M. (1982). *The meaning of educational change*. New York: Teachers College Press.

Glass, G. V, & Smith, M. (1977). *Pullout in compensatory education*. Boulder, CO: University of Colorado, Laboratory of Educational Research.

Hill, P. T., Wise, A. E., & Shapiro, L. A. (1989). *Educational progress: Cities mobilize to improve their schools* (R-3711-JSM/CSTP). Santa Monica: RAND.

Kimbrough, J., & Hill, P. T. (1981). *The aggregate effects of federal education programs*. Santa Monica, CA: RAND.

Leinhardt, G., Bickel, W., & Palley, A. (1982). Toward more effective remediation. *Teachers College Record, 84*(2), 391–422.

Levine, D. H. (1978). The social context of urban education. In D. A. Erickson & T. L. Reller (Eds.), *The principal in metropolitan schools* (pp. 106–129). Berkeley, CA: McCutchan.

March, J. G. (1978). American public school administration: A short analysis. *School Review, 86*, 217–50.

Meyer, J., & Rowan B. (1978). The structure of educational organizations. In J. W. Meyer & W. R. Scott (Eds.), *Environment and organizations* (pp. 79–109). San Francisco: Jossey-Bass.

Purkey, S. C., & Smith, M. S. (1983). Effective schools: A review. *Elementary School Journal, 83*, 427–454.

Slavin, R. E., Madden, N. A., Karweit, N. L. (1989). *Effective programs for students at risk*. Boston: Allyn & Bacon.

Venezky R. L., & Winfield, L. F. (1979). *Schools that succeed beyond expectations in teaching reading* (Tech. Rep. No. 1). Newark, DE: University of Delaware, College of Education. (Eric Document Reproduction Service No. ED 177 484)

Wacquant, J., & Wilson, W. (1989). The cost of racial and class exclusion in the inner city. *The Annals of the American Academy of Political and Social Sciences, 501*, 8–25.

Wimpleberg, R. K. (1989). The dilemma of instructional leadership and a central role for central office. In W. Greenfield (Ed.), *Instructional leadership: Concepts, issues, and controversies* (pp. 100–117). Boston: Allyn & Bacon.

Winfield, L. F. (1986a). Do Chapter I programs promote educational equity? A review and some comments. *Journal of Educational Equity and Leadership, 6*(1), 61–71.

Winfield, L. F. (1986b). Teacher beliefs towards academically at-risk students in inner urban schools. *The Urban Review, 18*(4), 253–267.

Winfield, L. F., & Stringfield, S. (1990). *A description of Chapter 1 schoolwide projects.* Baltimore, MD: The Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students.

Author

LINDA F. WINFIELD, Principal Research Scientist, Center for Research on Effective Schooling for Disadvantaged Students, The Johns Hopkins University, Baltimore, Maryland 21218. *Specializations:* social organization of schools, compensatory education.

Policy Issues: Chapter 1

Educational Evaluation and Policy Analysis
Winter, 1991, Vol. 13, No. 4, pp. 369–379

Modifying Chapter 1 Program Improvement Guidelines to Reward Appropriate Practices

Robert E. Slavin and **Nancy A. Madden**
Center for Research on Effective Schooling for Disadvantaged Students
The Johns Hopkins University

New accountability guidelines have helped to focus educators on the outcomes of Chapter 1 programs, but they may also be rewarding counterproductive practices. They may discourage early interventions, such as preschool, kindergarten, and first-grade programs, which increase the baseline for later gains. They may reward retentions, which significantly increase apparent normal curve equivalent (NCE) gains. They may focus teaching on narrow, easily measured objectives. This article proposes an alternative approach to Chapter 1 accountability which rewards schools for reducing the number of students who fail to meet minimum standards on broad-based, appropriate tests. Retained or untested students would be counted as not meeting minimum standards. Program improvement services would be greatly increased and made available to all Chapter 1 schools. Advantages and problems of this system are discussed.

While Chapter 1 and its predecessor, Title I, have always been service-delivery programs, they have also been accountability programs, requiring districts to evaluate and report the progress of Chapter 1 students on achievement tests. These accountability requirements have had a major impact on accountability procedures used by school districts for all students. The 1988 Hawkins-Stafford bill introduced new methods for evaluation of Chapter 1 programs and new roles for state and local education agencies tied to these evaluations. The changes are subsumed under the term *program improvement*. The intention of program improvement is to identify schools in which Chapter 1 students are not making adequate progress toward grade-level performance and to require these schools to reformulate their plans.

In concept, the idea of program improvement is a major step forward. For the first time, Chapter 1 is putting a major emphasis on the nature and quality of programs provided to children and the outcomes of these programs. The program improvement guidelines are surely identifying some schools which are, in fact, doing a poor job with low-achieving children and are giving them both an incentive to change and some assistance in doing so. Program improvement is also giving state departments of education more of a role in assuring program quality, as opposed to a primary emphasis on fiscal and regulatory monitoring (see Plunkett, 1991, this issue). Yet the very importance of the new program improvement guidelines places an added responsibility on them to be certain that they are fair and valid, and most important, that they reward schools for appropriate policies and practices. The purpose of this article is (a) to examine key aspects of program improvement guidelines to attempt to determine the degree to which they are likely to promote positive changes in school policies and practices and (b) to propose an alternative system designed to avoid the problems in the current one.

The identification of schools as being in need of program improvement is almost always based on the calculation of gains in nor-

Copyright 1991 by the American Educational Research Association. Reprinted by permission of the publisher.

mal curve equivalents (NCEs) from spring to spring or from fall to fall.[1] In principle, students who make the same progress as a test's norming population will receive the same NCE score each year (just as they would the same percentile rank). Most states have set a criterion for success of an average gain of more than zero and as many as three NCEs (see Heid, 1991, this issue) in each school, on the principle (laid out in the federal regulations) that Chapter 1 students, who are by definition performing below grade level, should be *gaining* on the national norming group to head toward grade-level performance.

NCE gains are computed from a baseline established from testing at the end of the previous grade. For example, a student who scores at an NCE of 30 in spring of first grade and 30 at the end of second grade would be considered to have made no gain; less than this is referred to as "negative gain." Of course, students whose scores are at the same NCE each year have gained in achievement, but they have not gained in comparison with the test's norming population.

In addition to NCE gains, states and districts are encouraged to include standards for "desired outcomes," such as reduced retentions, increased parent participation, or improved early childhood outcomes. However, because failure on any one of these can place a school in program improvement, most states and districts have either avoided "desired outcomes" or have made them easy to achieve (see Martinez, 1991).

In theory, the program improvement standards are sensible, in that they focus schools on the outcomes of instruction, not only on compliance with regulations regarding program operation. They also avoid the well-documented problems of fall-to-spring testing which plagued earlier Chapter 1 evaluation procedures. In practice, the new standards provide incentives for schools to improve their Chapter 1 programs, but they also create a few incentives which run counter to the intentions of the law and to any standard of common sense.

Not surprisingly, schools and school districts regard identification for program improvement as bad news. In school districts in which Chapter 1 plays a major role, many principals feel that identification of their school as being in need of program improvement will impact negatively on their careers, and in some districts this link is made explicitly by district administration. Clearly, most schools will be motivated to avoid being identified as in need of program improvement.

Most Chapter 1 schools probably try to avoid being identified for program improvement by attempting to improve the quality of their programs, as was intended in the legislation. However, the program improvement guidelines contain a few serious flaws which have the unintended effect of punishing schools for investing in early intervention (i.e., preschool, kindergarten, and first-grade programs) and rewarding them for retaining students and teaching a narrow set of skills. This article discusses these and other flaws in program improvement guidelines and proposes alternatives which would retain the positive features of the approach while eliminating these counterproductive features.

Flaws in Program Improvement Guidelines

Punishing Early Intervention

One of the most important themes raised by reformers of Chapter 1 is the idea that compensatory services should shift from an emphasis on remediation to an emphasis on prevention and early intervention (see, for example, Slavin, Karweit, & Madden, 1989). In recent years Chapter 1 dollars have increasingly been used to provide preschool, extended-day kindergarten, or intensive intervention in first grades for at-risk students, on the theory that it makes more sense to see that students begin with and maintain success than to let them fall behind in basic skills and only then provide remedial services. Programs such as Reading Recovery (Pinnell, 1989) and Success for All (Slavin, Madden, Karweit, Livermon, & Dolan, 1990), both of which provide one-to-one tutoring to at-risk first graders, have been highly effective in ensuring adequate reading skills among at-risk first graders, gains which have been maintained in later grades.

Yet program improvement standards may inadvertently punish schools for investing in any kind of early intervention; in fact, the more effective the early intervention, the more the school may be punished. The problem is that NCE gains are measured from a baseline established at the end of first grade. If students score very well at the end of first grade, this may make it more difficult to show continued NCE gains in Grades 2 and beyond. That is, a school which invests in preschool, kindergarten, or first-grade intervention may increase first-grade scores and therefore undermine its own program by having limited gains in the later grades. Regardless of whether or not schools continue to show gains in the later years, the impact of their early intervention will not be seen in the scores that matter most to Chapter 1 administrators.

To illustrate this, consider two identical schools with identical historical distributions of student scores. One school, Lowenslow Elementary, provides no Chapter 1 services until the second grade, at which time it offers traditional pullout services. The other, Brighton-Early Elementary, invests in extended-day kindergarten and one-to-one tutoring for the lowest-achieving first graders. Imagine that the effect of these interventions is to raise the achievement of at-risk students at Brighton-Early by 75% of a standard deviation, equivalent to an NCE gain of 16. Gains of this size are typical of studies of early intervention programs such as Reading Recovery (Pinnell, 1989) and Success for All (Madden, Slavin, Karweit, Dolan, & Wasik, 1991).

Table 1 shows a hypothetical distribution of scores for at-risk students at the two schools at the beginning of kindergarten (assume for the sake of argument that NCEs could be reliably measured at this grade level). "At-risk" is defined in this case as performing below an NCE of 30, the criterion for Chapter 1 services in both schools. Before any intervention, there are 20 kindergartners in each school scoring below 30 NCEs.

By the end of first grade, Lowenslow students have received no intervention, so they have stayed at the same performance level. In contrast, the 16 NCEs gained by Brighton-Early students have pushed all but six of the students over the criterion for Chapter 1 services. The NCE mean for all 20 students is now 35.3, but for the six remaining Chapter 1-eligible students it is 23.8 (see Table 1). This is now the baseline for NCE gains in Grades 2 and beyond. If at-risk students at Lowenslow and Brighton-Early then make gains of 3 NCEs each year, they will look (to Chapter 1 evaluations) like equally effective schools. Yet obviously they are not. At best, Brighton-Early's investment in early intervention does nothing to help it look good in the most important Chapter 1 evaluations (NCE gains in Grades 2 and up) and may increase its chances of ending up in program improvement by raising its end-of-first-grade baseline. The Chapter 1 assessment guidelines would fail to take notice of the most remarkable achievement of Brighton-Early's program: the fact that it dramatically increased end-of-first-grade performance and substantially reduced the number of students in need of remedial services.

The Chapter 1 Policy Manual does allow schools to submit data other than standardized test scores to evaluate preschool, kindergarten, and first-grade programs, but it forbids averaging any of these measures with those given in Grades 2 and up. This means that a school could have a very effective early intervention program (and have data to support it) but could still be identified for program improvement (on the basis of the standardized test scores from the higher grades).

The rationale for using end-of-first-grade scores as a baseline solved one problem inherent in earlier Chapter 1 evaluation systems, the relatively low reliability of end-of-kindergarten or early first-grade scores. However, flawed as it was, the earlier system at least communicated to schools that first-grade progress would count toward their success as a Chapter 1 school. In our own Success for All program (Madden et al., 1991), at least one principal moved tutors from the first to the second grade solely on the basis of the new Chapter 1 standards. It makes no instructional sense to allow at-risk students to fail in reading in first grade and then tutor them in second, but this was one unintended impact of the new guidelines. The principal was simply responding to a perception that first-grade achievement did not contribute to

TABLE 1
Hypothetical NCEs for Chapter 1-Eligible Students

	Beginning kindergarten		End of Grade 1		End of Grade 2	
Measure	Lowenslow	Brighton-Early	Lowenslow	Brighton-Early	Lowenslow	Brighton-Early
	30	30	30	(46)	(33)	((49))
	30	30	30	(46)	(33)	((49))
	29	29	29	(45)	(32)	((48))
	28	28	28	(44)	(31)	((47))
	27	27	27	(43)	30	((46))
	26	26	26	(42)	29	((45))
	25	25	25	(41)	28	((44))
	24	24	24	(40)	27	((43))
	23	23	23	(39)	26	((42))
	22	22	22	(38)	25	((41))
	21	21	21	(37)	24	((40))
	20	20	20	(36)	23	((39))
	18	18	18	(34)	21	((37))
	16	16	16	(32)	19	((35))
	14	14	14	30	17	(33)
	12	12	12	28	15	(31)
	9	9	9	25	12	28
	7	7	7	23	10	26
	4	4	4	20	7	23
	1	1	1	17	4	20
Mean NCE (A)	19.3	19.3	19.3	35.3	22.3	38.3
Mean NCE (B)			19.3	23.8	22.3	26.8
No. eligible	20	20	20	6	16	4
NCE gain, Grades 1–2					+3.0	+3.0
NCE gain from kindergarten			0	+16.0	+3.0	+19.0

Note. Mean NCE (A) = mean NCE (normal curve equivalent) for the original group of at-risk students (N = 20 in each school). Mean NCE (B) = mean NCE for students who still fall below an NCE of 30 at the end of first grade. Mean NCE (B) in spring of second grade minus mean NCE (B) in spring of first grade is the NCE gain for second grade. Scores with single parentheses are for students who were below an NCE of 30 in the prior year but not the current year. (These scores are used as posttests for NCE gains but not as pretests for the next year's gains.) Scores with double parentheses are for students who were in the original set of at-risk students but did not fall below an NCE of 30 in the prior or current year and therefore are not included in NCE-gain calculations.

the success of her school according to the new standards.

Rewarding Retention

A serious consequence of program improvement procedures is that they can inadvertently reward retention of students in Grades 2 and up. The reason for this is that when students take a test one year and then take the same test in the same grade the next year, their increase in percentile rank (and therefore NCEs) is very large, even though the students may in fact have gained little beyond the gain attributable to being a year older.

For example, consider a fourth grader who scores at the 20th percentile on the California Achievement Test (CAT, Form C) Total Reading Scale at the end of fourth grade and again at the 20th percentile on the CAT at the end of fifth grade. If the student had instead been retained, the same scale score would have placed him or her at the 43rd percentile. This apparent "gain" (from the 20th to the 43rd percentile) is entirely due to retention, not to improved performance. Put another way, this student went from a grade equivalent of 3.4 to one of 4.0. To have actually scored at the 43rd percentile in the fifth grade, the student would have had to increase to a grade equivalent of 5.3. That is, being retained gave the student an apparent bonus of 1.3 years of learning! One study which followed students from Grades 2–4 found that while students who were promoted each year gained an average of 5.4 NCEs, those who failed a grade gained 20.7, a "retention bonus" of 15.3 points (Karweit, 1991). Table 2 shows gains in percentiles and NCEs for retained students in a Florida study. The gains of nearly 17 NCEs in reading and 21 in math are far greater than the national average of 3.0 and 4.3 NCEs, respectively (Sinclair & Gutman, 1990). Since retained students come almost entirely from the ranks of students eligible for Chapter 1 services, even relatively small differences between schools in retention rates can lead to substantial differences in NCE gains. As a result, schools are unintentionally rewarded by program improvement guidelines for having high retention rates.

To illustrate the potential impact of retentions, imagine that Lowenslow Elementary decided to retain students in Grades 2 and above scoring at or below the 10th percentile and that this created an apparent NCE bonus of 15 points for each retained student. The three retained students (a retention rate of only 3.3% if there are a total of 90 second graders) would increase the NCE gains for second graders from +3.0 to +5.25. Retaining all students with scores below an NCE of 20 would still fail only 7.8% of Lowenslow first graders yet would increase apparent NCE gains from +3.0 to +8.25. Note that the same retention policies would produce no retentions at Brighton-Early Elementary, and as a result, this school (with a 0% retention rate and only four students [4.4%] still qualifying for Chapter 1 services) would appear much less successful than its twin. In urban school districts where retention rates often approach 20%, school-to-school variations in retentions could be much more important than actual program effectiveness in determining which schools are selected for program improvement (see Karweit, 1991).

Some school districts may deal with retentions by excluding them from the analysis of NCE gains. This is discouraged by the Chap-

TABLE 2
NCE Gains Due to Retention

Grade and test	Year in which student was retained		Year in which grade was repeated		NCE gain
	Percentile	NCE	Percentile	NCE	
Reading comprehension					
1	23	35	62	55	+20
2	12	25	42	46	+21
3	11	24	33	41	+17
4	11	24	28	38	+14
5	10	23	23	35	+12
Mean, 1–5					+16.8
Math computations					
1	20	32	66	59	+27
2	24	35	65	58	+23
3	31	40	69	61	+21
4	23	35	50	50	+15
5	24	35	55	53	+18
Mean, 1–5					+20.8

Note. NCE = normal curve equivalent score. Data are from Pinellas County, Florida, 1978–1979 (from Elligett & Tocco, 1983).

ter 1 Policy Manual (U.S. Department of Education, 1990), which emphasizes the requirement to seek a score for every child, but in any case it does not solve the problem, because retention still removes from the sample students who are by definition the lowest achievers.

The Chapter 1 Policy Manual does recommend that schools submit additional information, including retention rates, as a part of its documentation of program impacts, but there is no federal requirement to do so. In any case, a school which reports a low retention rate and therefore makes small NCE gains may still be in program improvement, whereas one which produced high apparent NCE gains by retaining large numbers of students will avoid program improvement (and need not report its retentions).

To the degree that program improvement guidelines accelerate a trend toward increasing retentions in elementary schools, they could have a disastrous effect on at-risk children. Long-term effects of retention are negative on many outcomes, academic as well as social and behavioral (Shepard & Smith, 1989). Disadvantaged students who have been retained before third grade are very unlikely to graduate from high school (Lloyd, 1978). Because of a realization of these long-term impacts, many urban school districts are now seeking to reduce retention rates which in many cases have exceeded 20% for certain grades (such as first). It would be tragic if Chapter 1 program improvement guidelines were to unwittingly punish districts for moving in this direction.

We are not suggesting that principals would deliberately fail more children to artificially increase NCE gains. However, schools which reduce their retention rates will have lower NCE gains and may be mistakenly singled out for criticism. Those which increase their retentions will have higher gains and may be mistakenly singled out for praise. This could ultimately result in a shift toward policies which increase retentions.

Rewarding Teaching the Test

A problem that has always plagued the use of standardized test scores for high-stakes accountability is the degree to which schools can appear to do well by teaching very narrowly those skills assessed on the tests (and ignoring other content). The new program improvement guidelines improve on this by emphasizing scores on advanced, as well as basic, skills, but the "advanced" skills in questions are really the same scales that have typically been given in the past, such as reading comprehension and math concepts and applications.

Practices which fall under the heading of "teaching the test" range from the relatively benign to the unethical (see Stringfield, Hartman, Pechman, & Brooks, 1985). At the benign end is "curriculum alignment," a focus of teaching efforts on the general skills or concepts being tested. Curriculum alignment is justifiable to the extent that one accepts what is on the test as the full range of what children should learn, an assumption that is perhaps tenable in some areas (such as math computation) and untenable in others (such as language arts tests without writing samples.) Teaching general test-taking skills also certainly falls on the benign end. However, both curriculum alignment and test skills are often overdone in high-stakes testing. For example, many urban elementary schools have little serious instruction in social studies, science, or writing, at least in part because these are not on the standardized tests. Many districts and schools have carefully examined the standardized tests and rooted out from their curriculum any objectives not explicitly tested. This has led in many cases to a great deal of teaching of isolated skills that is counterproductive to learning (but does improve test scores).

At the unethical end of the "teaching the test" continuum fall a variety of undesirable practices. In one common situation teachers become familiar with particular tests and make sure that they teach specific items known to be on the tests. For example, elementary vocabulary scales rarely involve more than 15 words, and teachers often learn these words and make certain to emphasize them in their teaching (in lieu of other kinds of vocabulary teaching).

The effects of teaching the test, teaching test-taking skills, and other means of increasing student scores without increasing their learning can be considerable. In a recent study, Koretz, Linn, Dunbar, and Shepard (1991) administered additional tests after standardized testing in "high-stakes testing" districts. One of the additional tests had been equated with the standardized tests in low-stakes districts, yet in the high-stakes districts school means on the alternative tests were substantially lower (by as much as 16 percentile points) than scores on the standardized tests used for accountability. The difference, the authors argue, is due to "teaching the test" and teaching test-taking skills; item-by-item comparisons (Flexner, 1991) and teacher surveys (Shepard & Dougherty, 1991) support this interpretation.

The new program improvement guidelines do not break any new ground in rewarding teaching to the test; all accountability programs suffer from this. What is new, however, is that the new standards raise the stakes for Chapter 1 schools and may thereby perpetuate a long-standing problem.

Other Problems of Chapter 1 Assessment

In addition to those mentioned above, there are several other problems with Chapter 1 assessment procedures which may not reward inappropriate policies but still may lead to problems in accurately identifying effective and ineffective programs. Chapter 1 assessments may be based on fewer than half of Chapter 1-eligible students (Bushner, 1991). If the students who took both tests were representative of all Chapter 1 students, this would be a minor problem, but it is more likely that missing students would make lower-than-average gains; thus, excluding them may overstate apparent gains. Worse, some schools may be less than relentless in obtaining a test from absent children who are unlikely to do well.

The reliability of NCE gains as an indicator of school effectiveness is another serious problem. Gain scores always have less reliability than do point-in-time scores, but this problem is compounded by any number of random factors, including the problem of missing students mentioned earlier. As one indication of the unreliability of NCE gains, Bushner (1991) compared fall-to-fall and spring-to-spring scores for the same schools in the same year and found no correspondence between the two; in fact, among the six schools followed, the highest-gaining school in the fall-to-fall assessment was the second to lowest in the spring-to-spring data, and the lowest-gaining fall-to-fall school was the second highest-gaining in spring-to-spring assessments. Consistency over time was also low; the second and third highest-gaining schools in the spring-to-spring 1988–1989 assessments were the two lowest-gaining schools on spring-to-spring 1989–1990 assessments. If NCE gains are to be used as high-stakes indicators of program effectiveness, they must be stable, meaningful, and reliable indicators. Clearly, this is not the case.

Finally, there is a problem of statistical regression that has long been noted in Chapter 1 evaluations (see Gabriel et al., 1985). That is, entirely because of random variation (e.g., bad luck), some students score below the cutoff for Chapter 1 services. Since bad luck is unlikely to happen twice, the student's score next year is likely to be higher, a situation which creates an apparent positive effect in the current Chapter 1 assessment system. In addition, fluctuation around the cutoff score can create an illusion of gain. For example, imagine that a student's NCE scores are 28, 32, 28, 32, 28 at the end of Grades 1–5, respectively, in a district using an NCE of 30 as a criterion for Chapter 1 eligibility. This student would show a gain of 4 NCEs in Grade 2. In Grade 3 he or she would not receive services, so the loss of 4 NCEs does not count on the Chapter 1 assessment. Then he or she appears to gain again in Grade 4, and so on. The Chapter 1 policy manual recognizes this problem and invites districts to correct scores for regression if they wish, but it is doubtful than any would do so because it is difficult and would have the effect of reducing scores.

Alternative Approaches

It is easy to criticize any accountability program but far more difficult to suggest practical alternatives. It is neither politically possible nor desirable to do away with accountability in Chapter 1 programs; we must have some outcome-based criterion on which to judge the impact of Chapter 1 in each school and district. The program improvement guidelines implemented under the Hawkins-Stafford bill are improvements in many ways over earlier procedures. Yet they still reward some policies we would want to discourage, and they punish policies we would want to encourage; thus, they are in need of major change. The following sections discuss a set of recommendations for a system which might accomplish the goals for which program improvement guidelines were originally designed.

Broad-Based Tests

The most important thing we have learned after 15 years of "accountability" in education is that high-stakes assessments do in fact drive instruction and other school practices but that if schools can find an easier way to affect assessments than to do a better job of teaching, they will often do so (Koretz et al., 1991). Therefore, assessments must be designed to be so broad and so appropriate to what we want students to do that they are worth teaching to and cannot be influenced by any kind of narrow teaching or "test-wiseness." The best model we have for such a test is NAEP, which uses matrix sampling, whereby different students take different portions of a very comprehensive test.

Broad-based tests should include some forms of "authentic" assessment, such as individually administered tests involving reading and comprehending real children's literature, writing samples, and open-ended problem solving in math, along with basic skills. The use of matrix sampling does not provide ideal student scores, but it does provide excellent information on school effectiveness. Most important, the use of a broad-based test would reward broad-based teaching, and use of "authentic" measures would reward the teaching of meaningful reading (not only skills), meaningful writing (not only language mechanics), and meaningful math (not only algorithms). Students might still take diagnostic tests to determine eligibility for services and as formative tests for school use, but the assessment of the program (as distinct from the students) would be based on these broader assessments. Some core tests might be given in every grade, while others (possibly including tests of science and social studies) might be given every few years.

An Alternative Model of Chapter 1 Assessment

We propose a system in which Chapter 1 schools are evaluated on the degree to which they can reduce the number of students in need of remedial services. In this system, states (or districts) would set minimum performance criteria for students at each grade level from pre-K on. The tests used at different grade levels could be different. For example, tests of preschool and kindergarten programs might focus on language development, and first-grade tests might include individual reading assessments, perhaps administered by Chapter 1 teachers from other schools. In the early years of such a system, passing scores could be established for existing standardized tests, but as states introduce new, more appropriate measures at selected grade levels, minimum performance criteria could be established for them. If matrix sampling is used, passing scores for each test form could easily be established. Existing standardized tests could be used until better tests are established at each grade level. Each student would be identified as meeting or not meeting minimum standards. Any students who were retained would be counted as not meeting standards, as would any student on roll in the spring and in the school all year who did not take a test. The idea here is to encourage schools to promote students and to try to obtain a valid test from every Chapter 1 student. Retaining students or failing to give them a test would provide no benefit to the school's scores because these students are counted into the school's total as not meeting standards.

In this system, the school would be rewarded for successively reducing its proportion of students failing to meet minimum standards, combining across all grade levels at which Chapter 1 dollars are spent. The initial baseline for this comparison would be based on a determination of how many students would have met minimum standards on tests given for three years before the new assessment system was implemented. That

282

is, standards for existing standardized tests would be established and used retrospectively to establish a baseline. This conforms to the critical principle that any baseline established for high-stakes assessment itself be a high-stakes assessment, so that schools would already be doing their best. After the first year, the proportion of students meeting minimum standards would always be compared with the proportion in the previous three years, so zigs and zags in baselines would not influence ratings of school success.

There are several important advantages to this system. First, it would allow for easy pooling of results across all grade levels; thus, schools could appropriately assess preschool, kindergarten, and first-grade students and have their successes added to the school's success. This also allows schools to use "authentic" tests, criterion-referenced tests, and other measures at some or all grade levels and thereby releases them from the requirement to use standardized tests solely because they produce NCE scores.

Because this system would use a point-in-time measure and because it would include an incentive to obtain a valid test from every student, the problems of missing data would be greatly reduced.

The system we have proposed would encourage early intervention. A child who received, for example, one-to-one tutoring in reading in first grade and therefore never needed further remediation would count every year as exceeding minimum standards. In the example comparing Brighton-Early Elementary to Lowenslow Elementary, the success of the early intervention at Brighton-Early would be clearly shown in this assessment system. What is important about Brighton-Early is not the increase in NCEs it brought about among its at-risk students but the reduction in the number of students falling below minimum standards while a zero retention rate was maintained.

Rewarding Success

Accountability systems primarily motivate educators to do their best by giving them internal benchmarks to judge their progress toward desired goals and by publicizing schools that are doing well or poorly. A principal wants his or her school to do well out of a sense of professional pride and wants to avoid falling into program improvement because it is embarrassing and potentially damaging to his or her career, not because the school staff has to change plans or attend workshops. Simply providing feedback on progress toward reducing the need for remedial services may be enough in most districts.

However, there may be a rationale for rewarding schools for increasing the proportion of students meeting minimum standards by giving the staff greater freedom in using Chapter 1 dollars and for imposing more restrictions on schools failing to reduce their Chapter 1 caseloads. For example, schools with a record of moving students out of Chapter 1 eligibility might qualify for schoolwide status, even if they do not meet the 75% poverty criterion. In contrast, schools failing to move students might be placed under substantial scrutiny by local and state regulators.

Potential Problems and Potential Solutions

As in any accountability system, there are several problems with the one we have proposed. First, a school undergoing major demographic changes might appear to be declining in the percentage of students meeting minimum standards. This could be dealt with by allowing schools to submit demographic data (e.g., increases in the percentage of students qualifying for free lunch) to explain any declines.

Second, a school which has always done a good job will start with a higher baseline than one that has done poorly and may therefore have more difficulty making further gains. Using a three-year baseline would somewhat diminish this problem, but the only real response is to note that any Chapter 1 school can always get better.

Third, it may be unfair to hold schools fully responsible for students new to the school. This problem might be solved by counting only students in the school for at least two years.

Another potential problem involves the use of a single criterion of success for each assessment. If this criterion were set too low, it might focus schools on minimum skills, whereas a high standard might lead them to focus on students near the passing score, ignoring those felt to be unlikely to pass under any circumstances. One solution to this problem might be to set two standards for each test: a "minimum" standard and a "basic" standard, where the basic level would roughly correspond to what is called "at grade level" on today's tests. Chapter 1 schools might be evaluated according to the degree to which they could move students beyond both of these standards. Advanced levels might also be established, particularly for use in schoolwide projects in which Chapter 1 has a legitimate interest in the performance of all students, not only low achievers.

Another important limitation of this system is that it uses as a criterion of success the very measure that qualifies a school for Chapter 1 funding in some districts; thus, a school doing well on the assessments could be reducing its Chapter 1 resources. A solution to this would be to base Chapter 1 funding solely on poverty. Other "hold harmless" provisions might be applied to make certain that schools which are reducing their Chapter 1 caseloads are not penalized for doing so.

Program Improvement

To live up to its name, program improvement must go beyond being primarily an accountability program and must devote much more attention and resources to actually improving programs. Chapter 1 needs to play a far greater role in staff development and in providing proven programs to students. The Hawkins-Stafford bill provided very modest funds for staff development, but a far greater focus on this aspect of Chapter 1 services is still needed (see Slavin, 1991). Ideally, schools should be able to receive on-site assistance to help them implement effective practices. This could be provided by state or regional Chapter 1 Effectiveness Centers staffed by professionals trained in various effective models and in the dissemination and implementation of effective practices (see Slavin, 1987). Such services should be available to all Chapter 1 schools or perhaps to schools serving large numbers of Chapter 1 students; these services should not be seen as a trip to the woodshed for schools who don't measure up. However, there would obviously be pressure for schools not meeting adequate standards to change programs and to invite in experts on effective models.

Conclusions

Throughout this article we have discussed the possibility that certain features of program improvement guidelines may reward schools for implementing inappropriate policies, such as avoiding early intervention, increasing retentions, or teaching to a narrow set of objectives. We do not mean to suggest that large numbers of principals would take advantage of these provisions to increase students' scores without increasing their learning. Rather, the danger is that schools which are working in good faith to implement early intervention models, to reduce retentions, and to encourage teachers to teach a full and appropriate curriculum may end up looking mediocre or worse in terms of NCE gain, even if their Chapter 1 children are in fact succeeding. At the same time, schools which are emphasizing remediation rather than prevention, retaining large numbers of students, and teaching narrowly to the standardized tests may mistakenly be held up as positive examples because of high NCE gains. If this occurs, innovative schools could become discouraged with reform and could return to the more traditional Chapter 1 practices which are more in line with the existing standards.

The solutions proposed in this article represent only a few among many possible ways we might revise program improvement guidelines. Our intention is simply to begin a discussion about modifications in program improvement guidelines to put them firmly behind (or a least not in the way of) school policies likely to benefit Chapter 1 children. Chapter 1 means a great deal to our most vulnerable children. We cannot rest until we are sure that Chapter 1 dollars are buying the most effective programs possible and that Chapter 1 policies are rewarding school practices conducive to the success of all children.

Notes

We would like to thank Sam Stringfield, James Smith, Mary Jean LeTendre, Nancy Karweit, Steve Davidoff, and Rita Altman for their comments on earlier drafts of this paper. However, we take full responsibility for the opinions presented.

This paper was written under a grant from the Office of Educational Research and Improvement, U.S. Department of Education (No. OERI-R-117-R-90002). However, any opinions expressed are our own and do not represent OERI positions or policies.

[1] A normal curve equivalent is a statistic similar to a percentile which ranges from 1 to 99, with a mean of 50 and a standard deviation of approximately 21.

References

Bushner, D. E. (1991, April). *Fall-to-fall testing versus spring-to-spring testing: What is the impact on a local community's Chapter 1 evaluation?* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

Elligett, J. K., & Tocco, T. S. (1983). The promotion/retention policy in Pinellas County, Florida. *Phi Delta Kappan, 64,* 733–735.

Flexner, R. J. (1991, April). *Comparisons of student mathematics performance on standardized and alternative measures in high-stakes contexts.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

Gabriel, R. M., Anderson, B. L., Benson, G., Gordon, S., Hill, R., Pfannenstiel, J., & Stonehill, R. M. (1985). *Studying the sustained achievement of Chapter 1 students.* Washington, DC: U.S. Department of Education.

Heid, C. A. (1991). The dilemma of Chapter 1 program improvement. *Educational Evaluation and Policy Analysis, 13,* 394–398.

Karweit, N. L. (1991). *What works in the Baltimore city public schools' Chapter 1 program?* Baltimore, Md: Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students.

Koretz, D. M., Linn, R. L., Dunbar, S. B., & Shepard, L. A. (1991, April). *The effects of high-stakes testing on achievement: Preliminary findings about generalization across tests.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

Lloyd, D. N. (1978). Prediction of school failure from third-grade data. *Educational and Psychological Measurement, 38,* 1193–1200.

Madden, N. A., Slavin, R. E., Karweit, N. L., Dolan, L. J., & Wasik, B. A. (1991). Success for all. *Phi Delta Kappan, 72,* 593–599.

Martinez, C. (1991, April). *Policy and technical issues on the identification of schools for improvement of Chapter 1 programs.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

Pinnell, G. S. (1989). Reading Recovery: Helping at-risk children learn to read. *Elementary School Journal, 90,* 161–182.

Plunkett, V. R. L. (1991). The states' role in improving compensatory education. *Educational Evaluation and Policy Analysis, 13,* 339–344.

Shepard, L. A., & Dougherty, K. C. (1991, April). *Effects of high-stakes testing on instruction.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

Shepard, L. A., & Smith, M. L. (Eds.). (1989). *Flunking grades: Research and policies on retention.* New York: Falmer.

Sinclair, B., & Gutman, B. (1990). *A summary of state Chapter 1 participation and achievement information for 1987–88.* Washington, DC: Decision Resources Corp.

Slavin, R. E. (1987). Making Chapter 1 make a difference. *Phi Delta Kappan, 69,* 110–119.

Slavin, R. E. (1991). Chapter 1: A vision for the next quarter-century. *Phi Delta Kappan, 72,* 586–592.

Slavin, R. E., Karweit, N. L., & Madden, N. A. (1989). *Effective programs for students at risk.* Boston: Allyn & Bacon.

Slavin, R. E., Madden, N. A., Karweit, N. L., Livermon, B. J., & Dolan, L. (1990). Success for All: First-year outcomes of a comprehensive plan for reforming urban education. *American Educational Research Journal, 27,* 259–278.

Stringfield, S., Hartman, A. Pechman, E., & Brooks, C. (1985, April). *Morality play: Cheating by any other name.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

U.S. Department of Education. (1990). *Chapter 1 policy manual.* Washington, DC: Author.

Authors

ROBERT E. SLAVIN, Co-Director, Early and Elementary Program, Center for Research on Effective Schooling for Disadvantaged Students, The Johns Hopkins University, 3505 N. Charles Street, Baltimore, Maryland 21218. *Specializations:* cooperative learning, school organization, programs for students at risk, research review.

NANCY A. MADDEN, Principal Research Scientist, Center for Research on Effective Schooling for Disadvantaged Students, The Johns Hopkins University, 3505 N. Charles Street, Baltimore, Maryland 21218. *Specializations:* programs for students at risk, mainstreaming, cooperative learning.

Educational Evaluation and Policy Analysis
Winter, 1991, Vol. 13, No. 4, pp. 399–406

Chapter 1 Program Improvement: Cause for Cautious Optimism and a Call for Much More Research

Sam Stringfield
The Johns Hopkins University

Shelley H. Billig
RMC Research Corporation

Alan Davis
University of Colorado at Denver

The program improvement provisions of the Hawkins-Stafford Amendments to Chapter 1 rest on the optimistic premise that school-level accountability pressures directed at Chapter 1 will lead to higher academic achievement for educationally disadvantaged students. Although the legislation may be unrealistic in assuming that improvement is primarily an act of will, it correctly focuses on the school as the appropriate unit for change. Principals of over 200 schools identified for program improvement in three states were surveyed to determine local responses to the new provisions. Over two-thirds of responding schools had begun to implement programmatic changes. Fully 84% supported the legislative provisions. Research is called for to study the effects of the legislation and to provide additional options to low-performing schools.

The most important and most optimistic sections of the 1988 Hawkins-Stafford Amendments to Chapter 1 were those dealing with program improvement. Hawkins-Stafford clearly reaffirmed that Title I/Chapter 1 was to be an educational program, not merely a funding program. By focusing on program improvement, the authors of the legislation set a tone and academic direction which has permeated discussions of Chapter 1. By linking program improvement to each school's Chapter 1 evaluation, the program improvement requirements reawakened local educators to the potentially powerful links between evaluation data and programming options. These were among the directions the original authors of Title I intended and were among the connections between evaluation and instruction which Congress had sought for over 20 years. The program improvement sections bore the unmistakable optimism of reformers. On a technical front, the authors assumed that local evaluations could be conducted which would possess sufficient reliability, validity, and clarity to serve two purposes. Hawkins-Stafford stipulates that local evaluations will be used to target poorly performing schools and to guide program improvement.

More broadly, the legislation *assumes* that there exist sufficient research, practical wisdom, and professional will so that teachers, paraprofessionals, and administrators in thousands of local schools—assisted by their districts and state departments of education—can and *will* improve the quality and quantity of services to their most needy students. Given the number of negative reports on the state of American education which had poured forth in the preceding five years,

Copyright 1991 by the American Educational Research Association. Reprinted by permission of the publisher.

Hawkins-Stafford might seem extraordinarily optimistic. However, the Congress tempered its high hopes with a series of steps that schools which do not meet the high expectations of Hawkins-Stafford must undertake. A stick came with the carrot.

We believe some of that optimism was justified and that over the next several years prudent local and national action can result in a state of affairs in which today's optimism becomes tomorrow's fact. In this article we examine program improvement on a practical level, overview the challenges facing persons attempting program improvement, describe one study of the effects on local educators of participating in program improvement, and draw implications for the 1993 reauthorization of Chapter 1.

The Practical Workings of Chapter 1 Program Improvement

The 1988 reauthorization of Chapter 1 required states to establish "Committees of Practitioners" which would set and periodically examine minimum standards for local schools to use in demonstrating the effectiveness of their compensatory education programs. As Heid (1991, this issue) notes, most state committees have set a minimum standard of "more than zero NCE gains on norm-referenced achievement test (NRTs)." (An NCE is a normalized standard score matching the percentile distribution at values of 1, 50, and 99, with a mean of 50 and a standard deviation of 21.06.) As a practical matter, the states were attempting to declare that a child must show more gain than would statistically be expected without Chapter 1. States have allowed schools to set additional criteria beyond NRT gains, but the wording of the law has typically been interpreted as indicating that if a school does not achieve *all* of its goals, it must enter into program improvement. This interpretation has provided little incentive for local schools to place additional requirements on themselves, and most schools appear to be declaring the minimum goals on the minimum number of criteria.

If a school's Chapter 1 students do not demonstrate the gains the school declared it would achieve, the school is identified as needing program improvement. Hawkins-Stafford states that during the first school year after identification, the school must consult with parents and write a program improvement plan. The plan is approved by the local educational agency (LEA) governing board and submitted to the state. Regulations suggest that minor improvements be implemented immediately; they allow up to a year for the implementation of more major changes. If there has been no improvement during the following year, the school must enter into a joint planning agreement with the state education agency (SEA) and local education agency. This process repeats until the school shows achievement gains.

Challenges Facing the Program Improvement Initiative

In the initial stages of any worthy undertaking, more reasons to predict failure can be listed than reasons to predict success. Program improvement is currently passing through such a period. The challenges facing honest efforts at implementing the program improvement sections of Hawkins-Stafford include, but are not limited to, the following: The conditions facing disadvantaged children have deteriorated considerably over the last 11 years, and they have been compounded by the fact that there is a difficult-to-reverse gravitation of highly skilled teachers and administrators to schools serving the most advantaged students; the levels of coordination between regular and Chapter 1 programs have often been inadequate; statute implications were often not fully explained to those who are being held most accountable; the existing technologies for achieving the goals of Hawkins-Stafford are often not strong and just as often poorly disseminated; and there remain technical problems with Chapter 1 evaluation techniques.

Worsening conditions. The conditions facing an increasingly large percentage of America's families and their children have worsened over the last decade, and it appears that the numbers of children being raised in poverty will increase over the next decade (Natriello, McDill, & Pallas, 1990; Wilson, 1987). The educational significance of those statistics is probably best understood through careful case studies. Kotlowitz

(1991), for example, describes the lives of two youngsters growing up amid the drugs, shootings, and general societal collapse of Chicago's slums. The children live in abominable, unsafe conditions. They regularly have to hide from the stray bullets of drug- and gang-related shootouts. They do not always have enough food. Children in these circumstances can hardly be expected to concentrate on academics with the same single-minded ease as do suburban children, regardless of the quality of their schools.

"*The rich [schools] get richer and the poor. . . .*" There exists in American education a gravitational-like pull from schools offering the most resources to attract the most highly trained principals and staff. Wimpelberg, Teddlie, and Stringfield (1989) described this phenomenon as it affected two schools in one large system. The non-Chapter 1 magnet school had over 40 highly qualified applicants for every teaching opening, while the principal of the school serving a 90+% free-lunch population often had to wait months for one qualified applicant for a position. It is relatively easy to motivate teachers to implement new programs when they know that there are 40 other qualified teachers eagerly awaiting their departure. The same motivational task is much more difficult when a principal knows that if he or she pressures a teacher too much, that teacher may leave, and some of the school's students may be served by "permanent substitutes" for the remainder of the school year.

Historical isolation of categorical programs. In many school districts, Chapter 1 had inadvertently become an isolated categorical program. The long-standing federal requirement that programs "supplement not supplant" regular programs had heightened this isolation. Many regular classroom teachers felt no connection at all to "Chapter." Many regular and Chapter 1 teachers had little or no knowledge of each other's programs, curricula, or instructional techniques. Yet, the Hawkins-Stafford Amendments hold whole schools accountable for students' success. Regular teachers and principals were unaccustomed to having input into the design and evaluation of their Chapter 1 services.

Occasional staffing concerns. In too many school districts, "Chapter" had become the retreat of highly senior staff who, regardless of their instructional talents, no longer wished to deal with the demands of 20–40 students and who preferred working with groups of 2–5 students at a time in controlled environments. In some districts, Chapter 1 had become the last refuge for teachers who would have been placed on probation or fired if a convenient place had not been available "where they won't hurt as many children." The chief teachers' union representative of one of the nation's 50 largest school districts once explained to the first author that she was able to virtually eliminate disciplinary actions against teachers by having incompetent teachers shifted to Chapter 1. The school districts that allowed such practices now lack solid foundations on which to build improving programs.

Breakdowns in information flow. In many instances, the intentions and requirements of the legislation were not clearly articulated to local principals, teachers, and paraprofessionals who were responsible for successfully achieving provisions of the new law. In most states annual meetings of local Chapter 1 coordinators are held to introduce any new wrinkles in the Chapter 1 law and regulations. Many local Chapter 1 coordinators report learning about small changes which were mandated with much fanfare one year and retracted the next. This history tended to have a deadening effect on local program coordinators' reactions to announced changes at the federal level. When truly major change came, local coordinators tended to take a "wait and see" attitude toward the new law. Hawkins-Stafford was not a "wait and see" piece of legislation. What had been adaptive behavior for local federal program administrators became dysfunctional.

State Chapter 1 meetings are rarely attended by principals, and almost never by teachers. In probably thousands of school districts, the school-site principals, teachers, and paraprofessionals who were later held accountable for achievement gains under Hawkins-Stafford were often unaware of the requirement until they learned that they had been "targeted" (a most unfortunate choice of words) for program improvement. Not surprisingly, "targeted" people have tended to respond defensively. Defensiveness is an excellent negative predictor of meaningful instructional change. Many of the federal, state, and local implementors of Hawkins-Stafford inadvertently created worst-case scenarios for initiating meaningful program improvement.

Lack of research support. Program improvement assumes the availability of programs which are effective and transportable. Unfortunately, federal funding of programmatic research came to a virtual close after the unsuccessful Follow Through evaluations (Stallings & Kaskowitz, 1974). Good (personal communication, February, 1989) estimated that, in constant dollars, federal funding for educational research had dropped over 80% since 1973. To hold schools accountable for making programs work without providing research clearly indicating which programs achieve specific goals and which don't was a considerable act of optimism. Promising new beginnings in federal support for Chapter 1 research have been reported by Plisco and Scott (1991), but much more is needed.

The programs which have at least modest evidence of effectiveness have not been well disseminated. Fullan (1982, 1991), Rosenblum and Louis (1981), Louis and Miles (1990), and Showers, Joyce, and Bennett (1987) provide clear conclusions regarding the conditions necessary for meaningful implementation of new programs. These include multiyear processes with central roles for leadership and technical assistance. Clear visions and goals, early success, sustained interactions among the people being asked to change, and intensive staff development for everyone involved are research-supported elements of sustained change. At the teacher level, presentation of theory combined with modeling of appropriate new behaviors, opportunities to practice new behaviors, quick and accurate feedback, and ongoing coaching are all supported by research.

These are hardly cost-free elements. Yet funding for the National Diffusion Network and other dissemination activities shrank dramatically during the 1980s and now offers very few of the activities and almost none of the extended follow-up necessary for successful program implementation. The Council of Chief State School Officers (1991) estimates that there are 9,000 schools currently identified as needing program improvement. For 20 years state education agencies have been required to focus on technical compliance issues. Even if there were enough research available, there simply are not enough diffusion resources to meet this wellspring of demand.

Evaluation use issues. Finally, there are technical and substantive problems with measurement and evaluation in Chapter 1. Much has been written about this elsewhere (see, for example, Davis, 1991, this issue; and Slavin & Madden, 1991, this issue). If there is one chance in 20 that a program has been misidentified on a technicality, then there will be at least 15 of 20 schools who perceive themselves to be the one. This can lead to either of two further technical problems. Principals and faculty may become so convinced that they have been unfairly targeted that they resist all suggestions and efforts at program improvement. Alternately, schools may opt to teach "test taking skills" or simply teach the test. Such processes not only take time from instruction and risk invalidating the scores, but they also may greatly raise the "pretest" scores from which next year's "posttests" must show gains. Thus they risk creating and then perpetuating procedures which are, at best, invalidating the evaluation and, at worst, unethical.

In sum, the above difficulties could lead to an easy prediction of failure for Chapter 1 program improvement. Yet while adequate national studies of program improvement have not been funded, we are inclined to believe that in most schools and most states program improvement is working.

In a previous article, we reviewed four studies which indicated that schools which participate in a year-long guided process of planning and implementing Chapter 1 program improvement tend to begin showing achievement gain the following year (Stringfield, Billig, & Davis, 1991). One topic which was not central to those studies but which has become central to the ongoing debate over the Hawkins-Stafford Amendments concerns the actions and perceptions of staff members in schools targeted for program improvement. It is important to know what local teachers and principals are doing and thinking once they are identified for program improvement. If they respond defensively because they see no practical options for improving students' learning, then program improvement will fail. If most see genuine steps they can take to improve their programs and get themselves "untargeted," and they are taking those steps, then program improvement may succeed.

Local Educators' Perceptions of the Effects of Chapter 1 Program Improvement

To determine local educators' responses to being targeted for program improvement, questionnaires were sent to the principals in all schools identified for Chapter 1 program improvement in three states. One of the states is located in the South, one in the Midwest, and one in the Southwest. Over 200 questionnaires were sent, and responses were received from just over 52% of the schools surveyed. It is possible that this response rate resulted in a biased sample, but the direction of any bias is not clear to the researchers. The respondents were not always principals. In 15% of the responses, questionnaires were completed by Chapter 1 coordinators, teachers, curriculum developers, and other staff. In several cases, the entire school's improvement team completed copies of the questionnaire. When more than one respondent from a school answered a survey, the answers were combined so that all schools would receive equal weighting.

Questions were open ended, and many respondents provided multiple answers to single questions. This resulted in a rich data set, but one that did not lend itself to quantitative analyses. Results will be presented as they relate to six overriding issues: Did the participants understand why they had been identified? Had they actually made changes? What factors did educators perceive to be inhibiting change? Facilitating change? Was evidence of outcomes available? Finally, the questionnaire asked educators whether the program improvement legislation was having a positive or negative effect on schools and schooling. The first five questions will be addressed briefly, followed by a more detailed discussion of the sixth.

Did the participants understand why they had been identified? Yes. Over 90% of the respondents stated that their school had been identified because their Chapter 1 students had not shown sufficient gains on norm-referenced tests (NRTs). In a few cases, schools had volunteered for program improvement, reporting that they believed the process would be a healthy one for their school.

Had schools actually made changes? This question was complicated by issues of timing. The schools had been identified on the basis of 1989–1990 achievement data (the first year of implementation), and questionnaires were sent during the spring of 1991. Hawkins-Stafford requires that by the end of the first year after identification, schools produce a plan of action. Neither the law nor subsequent regulations require that schools fully implement their plans in year one.

Over two-thirds of the schools had begun implementation. The most frequent exceptions were schools preparing to move to computer-based provision of Chapter 1 services. At several of those sites, the schools were awaiting the delivery of hardware and software, and the provision of training. Among the schools which had begun implementation, no clear patterns were apparent from the responses. Although no single type of change predominates, it is clear that in this sample there was no wholesale movement toward either test-driven instruction or blaming the measurement tool for the purported lack of success. Some schools which had relied on pullout programs were moving to in-class, some in-class were moving to computer-assisted instruction (CAI), and other than those moving to CAI, none indicated a shift to a nationally recognized program model, such as Reading Recovery or Accelerated Schools.

What factors did educators perceive to be inhibiting or facilitating change? An interesting finding concerned the role of each state department of education in structuring the program improvement processes. In two of the three states, the state Chapter 1 directors had opted to use their regional Chapter 1 Technical Assistance Center (TAC) and/or Rural Technical Assistance Center (R-TAC) as part of a year-long program improvement planning process. This resulted in increased reports of collaboration among staffs and a wide variety of program changes. The third state's Chapter 1 director had opted to encourage individual schools and districts to use a self-assessment questionnaire based on the "Thirteen Attributes of Effective Compensatory Education Programs" (Griswold, Cotton, & Hansen, 1986). Literal interpretations of data gathered on that instrument have been found to suggest that "parent involvement" and "coordination among programs" are the two greatest needs in Chapter 1 programs (Davis & Billig, 1989). Not surprisingly, the majority of projects in the third state focused their improvement efforts on parent involvement and increased coordination.

The most commonly reported answer to a question on factors *inhibiting* change was "none." Second was a perceived lack of support. Some of the teacher respondents perceived that their principals were not open, concerned, or involved in the process. Some of the principals saw a rigid central office staff at the center of their problems. Others reported that the lack of funds, difficulties in scheduling, problems with physical space, and the resistance of some professionals were difficulties.

The most commonly reported answers to a question on factors *facilitating* change included the commitment of the whole school's faculty to school improvement, collaborative problem solving, and administrative support and monitoring.

Was evidence of outcomes available? Given that virtually all of the schools were in program improvement because of lack of gains on NRTs and that in most districts the questionnaire was received before the next year's test data were available, it was not surprising that most respondents reported that it was "too early to tell" whether improvements had been successful. A few schools were able to report achievement gains, and several schools noted that parents seemed more involved and more pleased with Chapter 1. Others noted an increase in staff involvement in and ownership of Chapter 1, and one reported students reading more widely as evidence of success.

Toward the end of the questionnaire, school personnel were asked, *"Do you think that the Chapter 1 program improvement legislation has had a positive or negative impact on Chapter 1 as a whole? Why?"* Fully 84% of the respondents stated that the program improvement statutes were having a positive effect on Chapter 1. Responses varied from identifying with the idealism of the legislation ("We feel that the impact of program improvement is positive because school improvement is a vital part of the educational process. We should strive for the best. . . . ") to a rather conservative pragmatism ("The legislation had a positive impact, making schools more accountable. It has really helped in this district because the ownership for the Chapter 1 program returned to the school level . . . "). Teachers and principals reported being more aware of the needs of students, the needs of their school, and of the options available to them.

Five percent stated that the effect of the legislation was negative. Persons expressing negative views included principals and teachers. The most frequent reason given for a negative response was a questioning of the validity or relevance of the NRTs. One Chapter 1 teacher stated a resentment that Chapter 1 teachers "are held totally accountable for a student's improvement, where the classroom teacher isn't."

Eleven percent of the respondents stated that the effect was neutral or that it was too early to tell. One principal observed that as long as control of Chapter 1 remained centralized at the district office, the program's potential for effect was fixed and the law's effects were insignificant.

Discussion and Implications

Much of the public debate regarding the program improvement sections of the Hawkins-Stafford Amendments concerns the assumption of negative perceptions resulting from being "targeted" for program improvement. Our study indicates that it is possible to structure program improvement processes so that the large majority of local educators, even those who have been targeted, are supportive of the program improvement goals and processes. If a potential stigma on adults can be turned into productive change, then the debate on Chapter 1 program improvement can move forward to more child-focused issues. This requires patience, wisdom, and firm but gentle guidance from several levels of bureaucracy above the teacher and school.

A second implication concerns staff development, coordination, and buy-in. Many regular classroom teachers and principals in Chapter 1 schools were unaware that their schools had become accountable for the measured academic growth of their poorest performing students. Most had not known what an NCE was, much less felt responsible for the production of more of them. Often they felt unconnected to the processes and outcomes of compensatory education. These are staff development problems. The current research gives some cause for optimism regarding the salutary effects of cross-program, often schoolwide staff development. Such staff development efforts can improve both coordination and regular classroom teachers' buy-in to the goals and processes of Chapter 1.

A third finding concerned the handling of program improvement by the states. The three states involved in our survey had used moderately differing program improvement processes and had reaped differing results. A very large natural experiment in change is happening in Chapter 1 today. It is important, but is not being thoroughly researched. Our results indicate that state-level differences in process may be producing considerable differences in local educators' perceptions of the law, local options, and the value of Chapter 1.

A fourth issue which emerged in our work with various states and in analyses of the data concerns Chapter 1 evaluations. The Volume 1, Number 2, of *Educational Evaluation and Policy Analysis* (1979) was devoted to technical problems surrounding the Title I Evaluation and Reporting System (TIERS), and little technical work has been conducted on the area since. This is in spite of the fact that significant breakthroughs have been made in the areas of testing and, importantly, in the technical requirements for measuring change (e.g., Raudenbush & Bryk, 1989; Rogosa, 1989; Willett, 1989). One clear implication of this later research is the need for "three points in time" for measuring gain. Requirements for "targeted" program improvement need to be based on the strongest possible evidence. Both technical issues and evaluation use issues in Chapter 1 need new attention.

There is a great need for more solid research on practical options for program improvement. It is not enough to tell programs that they must improve. If the federal government is to mandate change, then it should also provide a considerable list of previously researched and independently evaluated options for change. These should be accompanied with a matrix of conditions under which various programs might be more or less viable choices. The development of such knowledge would require a great deal of additional research. As the accounting firm of Arthur Anderson & Co. has observed, any industry which spends as small a percentage of its total operating budget on research as does education would soon find itself hopelessly outstripped by its competitors. Moreover, the logical level for funding of educational research is federal (Measelle & Egol, 1990).

If program improvement strategies are to be fully implemented, Congress will have to allocate much more money for state and local support of change. Under Hawkins-Stafford states receive $90,000 per year to facilitate program improvement. This is an inadequate amount for Wyoming's needs and hardly worth mentioning in a state the size of California. Studies of change consistently find that change takes time, coordination, leadership, and *multiyear support.*

We began this article with a long list of reasons why the program improvement requirements of the Hawkins-Stafford Amendments to Chapter 1 might have difficulty in being successfully implemented. The data from our and other studies indicate that carefully implemented program improvement can become a force for more fully integrated, thoughtful educational programming and higher achievements for disadvantaged students. The data from our modest sample also indicate that local educators can overcome concerns regarding being "targeted" for program improvement and can focus on the important issues of providing the best possible services to children. The primary challenge facing the authors of the next reauthorization of Chapter 1 will be in providing sensible policies, research, and funding to support enhanced program improvement.

References

Council of Chief State School Officers (1991). *Chapter 1 program improvement and innovation across the states—An overview and survey excerpts*. Washington, DC: Council of Chief State School Officers.

Davis, A. (1991). Upping the stakes: Using gain scores to judge local program effectiveness in Chapter 1. *Educational Evaluation and Policy Analysis, 13* (4), 380–388.

Davis, A., & Billig, S. (1989, April). *The Chapter 1 self-assessment instrument: The internal consistency of a program improvement tool.* Paper presented at the meeting of the American Educational Research Association, San Francisco.

Fullan, M. (1982). *The meaning of educational change*. New York: Teachers College Press.

Fullan, M. (1991). *The new meaning of educational change*. New York: Teachers College.

Griswold, P., Cotton, K., & Hansen, J. (1986). *Effective compensatory education sourcebook, Vol. I.* Washington, DC: U.S. Department of Education.

Heid, C. A. (1991). The dilemma of Chapter 1 program improvement. *Educational Evaluation and Policy Analysis, 13*(4), 394–398.

Kotlowitz, A. (1991). *There are no children here*. New York: Doubleday.

Louis, K., & Miles, M. (1990). *Improving the urban high school.* New York: Teachers College.

Measelle, R., & Egol, M. (1990). *A new system of education: World-class and customer-focused.* New York: Arthur Anderson & Co.

Natriello, G., McDill, E., & Pallas, A. (1990). *Schooling disadvantaged children: Racing against catastrophe*. New York: Teachers College.

Plisko, V., & Scott, E. (1991). Planned evaluations of Chapter 1. *Phi Delta Kappan*, 72(8), 590–591.

Raudenbush, S., & Bryk, T. (1989). Methodological advances in analyzing the effects of schools and classrooms on student learning. In E. Z. Rothkopf (Ed.), *Review of research in education* (Vol. 15, pp. 423–476). Washington, DC: American Educational Research Association.

Rogosa, D. (1989). Longitudinal analysis of student achievement data: Issues for Chapter 1 evaluation. In Policy Studies Associates, *Planning papers for the national longitudinal study of Chapter 1* (pp. 423–475). Washington, D.C.: Policy Studies Associates.

Rosenblum, S., & Louis, K. (1981). *Stability and change*. New York: Plenum Press.

Showers, B., Joyce, B., & Bennett, B. (1987, November). Synthesis of research on staff development: A framework for future study and a state-of-the-art analysis. *Educational Leadership, 45*(3), 77–87.

Slavin, R., & Madden, N. (1991). Modifying Chapter 1 program improvement guidelines to reward appropriate practices. *Educational Evaluation and Policy Analysis, 13*(4), 369–379.

Stallings, J., & Kaskowitz, D. (1974). *Follow Through classroom observation evaluation*. Menlo Park, CA: Stanford Research Institute.

Stringfield, S., Billig, S., & Davis, A. (1991). Implementing a research-based model of Chapter 1 program improvement. *Phi Delta Kappan*, 72(8), 600–606.

Willett, J. (1989). Questions and answers in the measurement of change. In E. Z. Rothkopf (Ed.), *Review of research in education* (Vol. 15, pp. 345–422). Washington, DC: American Educational Research Association.

Wilson, W. (1987). *The truly disadvantaged*. Chicago: University of Chicago.

Wimpelberg, R., Teddlie, C., & Stringfield, S. (1989). Sensitivity to context: The past and future of effective schools research. *Educational Administration Quarterly, 25*(1), 82–107.

Authors

SAM STRINGFIELD, Principal Research Scientist, Center for Research on Effective Schooling for Disadvantaged Students, The Johns Hopkins University, 3505 N. Charles Street, Baltimore, Maryland 21218. *Specializations:* school effects, Chapter 1 effects, education for disadvantaged children.

SHELLEY H. BILLIG, Director, Region 5 Rural Technical Assistance Center, RMC Research, 1512 Larimer Street, Suite 540, Denver, Colorado 80202. *Specializations:* school improvement, program evaluation, and Chapter 1.

ALAN DAVIS, Assistant Professor of Research and Evaluation Methodology, University of Colorado at Denver, School of Education, Campus Box 106, PO Box 173364, Denver, Colorado 80217. *Specializations:* education policy analysis and school improvement.

Journal Descriptions

American Journal of Education is devoted to original inquiries in education, to the evaluation and synthesis of educational scholarship, and to scholarly commentary on educational practice. The journal seeks (1) to bridge and integrate the intellectual, methodological, and substantive diversity of educational scholarship, and (2) to encourage a vigorous dialogue between educational scholars and practitioners about matters of policy and governance and the management and conduct of schools. The journal is hospitable to scholarly writing on a variety of educational topics, from differing conceptual, methodological, and substantive approaches. Four kinds of papers are especially encouraged: (1) research reports; (2) scholarly writing of other kinds—for example, theoretical statements and philosophical arguments; (3) critical syntheses of a field of educational inquiry; and (4) integrations of educational scholarship, policy, and practice. For additional information contact: Philip W. Jackson, Editor, American Journal of Education, 5835 Kimbark Avenue, Chicago, Illinois 60637.

American Sociological Review is the official journal of the American Sociological Association and the leading journal in sociology. It publishes original (not previously published) works of interest to the discipline in general, about new theoretical developments, results of research that advance our understanding of fundamental social processes, and important methodological innovations. All areas of sociology are welcome. Emphasis is on exceptional quality and general interest. For additional information contact: Paula S. England, Editor, American Sociological Review, Department of Sociology, University of Arizona, Tucson, Arizona 85721.

Educational Evaluation and Policy Analysis (EEPA) focuses on educational evaluation, educational policy analysis, and the relationship between the two activities. It strives to serve the multiple needs of the diverse specialists currently working in educational evaluation and policy analysis. EEPA deals not only with theoretical and methodological issues but also with the intensely practical concerns of individuals engaged in the evaluation of educational enterprises and the formulation of educational policy. It apprises readers of current developments in the emerging educational specializations of evaluation and policy analysis. For additional information contact: American Educational Research Association, Publications Department, 1230 17th Street NW, Washington, DC 20036-3078.

Journal of Educational Psychology publishes original investigations dealing with learning and cognition, social and emotional processes, and human development as they relate to problems of instruction. *Journal* articles pertain to all levels of education and to all age groups. For additional information, contact: Joel R. Levin, Editor, Department of Educational Psychology, University of Wisconsin, 1025 West Johnson Street, Madison, Wisconsin 53706–1796.

Journal Descriptions

Reading Research Quarterly is published four times a year by the International Reading Association as a service to its members and other interested persons, and is intended to provide a forum for the exchange of information and opinion on theory, research, and practice in reading. For additional information contact: Reading Research Quarterly, The Ohio State University, 210A Arps Hall, 1945 North High Street, Columbus, Ohio 43210-1177.

Review of Educational Research (RER) publishes critical, integrative reviews of research literature bearing on education. Such reviews should include conceptualizations, interpretations, and syntheses of literature and scholarly work in a field. RER encourages the submission of research relevant to education from any discipline, such as reviews of research in psychology, sociology, history, philosophy, political science, economics, computer science, statistics, anthropology, and biology, provided that the review bears on educational issues. RER does not publish original empirical research unless it is incorporated in a broader integrative review. RER will occasionally publish solicited, but carefully refereed, analytical reviews of special topics, particularly from disciplines infrequently represented. For additional information contact: American Educational Research Association, Publications Department, 1230 17th Street NW, Washington, DC 20036-3078.

Urban Education is a publication of Corwin Press, a subsidiary of Sage Publications, Inc. The journal, publishing refereed manuscripts quarterly, is the premier journal addressing issues related to the education of urban youth. Once a year, *Urban Education* offers a special issue focusing on one facet of education in the inner city. The journal not only looks at the problems in urban schools; it also offers solutions, in part by highlighting successful programs in the inner cities. In addition to publishing state-of-the-art research and conceptual articles in each issue, *Urban Education* regularly features essay reviews of current publications in the area of schooling in urban centers. For additional information contact: Dr. Kofi Lomotey, Editor, Urban Education, 111 Peabody Hall, College of Education, Louisiana State University, Baton Rouge, Louisiana, 70803.

☆ U.S. GOVERNMENT PRINTING OFFICE:1993-342-111/93153